EAS Journal of Parasitology and Infectious Diseases
Abbreviated Key Title: EAS J Parasitol Infect Dis
ISSN: 2663-0982 (Print) & ISSN: 2663-6727 (Online)
Published By East African Scholars Publisher, Kenya

OPEN ACCESS

Volume-8 | Issue-2 | Apr-Jun- 2026 | DOI: https://doi.org/10.36349/easjpid.2026.v08i02.002

Review Article

Role of Artificial Intelligence in Vector-Borne Disease Management in India

Ms. Navya Mall[1]*, Dr. Tridibesh Tripathy[2], Ms. Sanskriti Tripathy[3]

[1]Research Scholar for Ph.D. in Law in the Faculty of Law, University of Delhi, Delhi, LL.M. (Master of Laws, BHU, Varanasi), LL.B. (Bachelor of Laws, DU, Delhi)
[2]BHMS (Utkal University, Bhubaneswar), MD (BFUHS, Faridkot), MHA (TISS, Mumbai), Ph.D. in Health Systems Studies (TISS, Mumbai), Homoeopathic & Public Health Expert, Visiting Professor, Master of Public Health (Community Medicine) program, Department of Social Work, Lucknow University, Lucknow, UP, India
[3]Student of Final Semester, B.Tech in Biotechnology, Bennett University, Greater Noida, Uttar Pradesh

Article History
Received: 04.03.2026
Accepted: 30.04.2026
Published: 04.05.2026

Journal homepage:
https://www.easpublisher.com

Quick Response Code

Abstract: India carries one of the heaviest vector-borne disease burdens in the world. Dengue, malaria, chikungunya, japanese encephalitis, and kala-azar together kill thousands of Indians every year and push millions more into catastrophic health expenditure. The country's surveillance infrastructure, while improving steadily, still relies on passive case detection, delayed reporting, and manually aggregated data that reaches decision-makers too late to prevent outbreaks from spreading. Over the past decade, artificial intelligence has begun entering this space through disease forecasting models, satellite-driven risk mapping, social media-based early warning tools, and AI-assisted diagnostics. This paper examines these developments specifically in the Indian context. It surveys the technical tools being deployed, their performance as reported in Indian studies, the regulatory and ethical questions they raise under Indian law, and the institutional obstacles that explain why AI in Indian public health has so far produced more pilot projects than operational systems. The argument advanced here is that AI can genuinely improve India's vector-borne disease response, but only if it is built on better data infrastructure, governed under a coherent regulatory framework, and deployed with the realities of India's district health system in mind rather than the requirements of a research paper.

Keywords: AI, Vector, NTD, NVBDCP, Forecasting, WHO, NIHFW.

INTRODUCTION

The concept of vector borne diseases comes under the domain of infectious epidemiology. Vector is defined as an arthropod or any living carrier like snail that transports an infectious agent to a susceptible individual. Transmission is through mechanical & biological ways. These diseases are clubbed under Neglected Tropical Diseases (NTD) by the World Health Organization (WHO). NTDs constitute 21 diverse infectious conditions that include parasitic, bacterial, viral & fungal diseases [1, 2].

India reported approximately 193,000 confirmed dengue cases in 2022, with the actual number almost certainly five to ten times higher once under-reporting is accounted for. Malaria resulted in 83 confirmed deaths recorded by the National Centre for Vector Borne Diseases Control that same year, though modelled estimates from the Global Burden of Disease Study place the true mortality figure closer to 9,400 reflecting the chronic shortfall between passive surveillance data and actual disease burden. Japanese encephalitis continues to claim children's lives in eastern Uttar Pradesh and Bihar every monsoon season. Kala-azar persists in Bihar and Jharkhand despite years of elimination campaigns. Chikungunya re-emerges periodically in urban centres with little warning. The common thread across all of these diseases is the same: by the time the outbreak is confirmed, the window for early intervention has largely closed [3-5].

India's vector-borne disease problem is not simply a question of healthcare capacity, though capacity matters enormously. It is fundamentally a surveillance

***Corresponding Author:** Ms. Navya Mall
Research Scholar for Ph.D. in Law in the Faculty of Law, University of Delhi, Delhi, LL.M. (Master of Laws, BHU, Varanasi), LL.B. (Bachelor of Laws, DU, Delhi)

problem. The country's Integrated Disease Surveillance Programme, launched in 2004 and progressively expanded since, has built real-time reporting infrastructure at district level across most states. But reporting is only as good as the data entering the system. Most cases of dengue in India are managed at home or in private clinics that have no obligation to report to public health authorities. Laboratory-confirmed diagnosis is concentrated in government hospitals that serve a minority of patients in most districts. The result is a system that detects the tip of the iceberg and tries to draw epidemiological conclusions from it [3-5].

This is where artificial intelligence offers something concrete. Machine learning models can extract outbreak signals from proxy data sources, weather station readings, satellite imagery, internet search trends, and social media posts , that are available in near-real time and do not depend on the diagnostic and reporting pathways that conventional surveillance requires. They can forecast dengue incidence weeks in advance with accuracy sufficient to trigger operational responses. They can identify the specific urban neighbourhoods or rural blocks where mosquito breeding is likely to be highest at a given point in the season. They can help peripheral health workers diagnose malaria accurately without a trained microscopist on site [6].

Whether they actually do these things reliably in India is a more complicated question. The research literature is encouraging but uneven. Most published studies report model performance on historical data from a single city or a handful of districts, under conditions that are somewhat more favourable than what a district health officer faces when trying to use the tool in the field. The gap between research accuracy and operational utility is real, and it has not been honestly reckoned with in the policy discourse around AI and health in India. This paper attempts that reckoning not to dismiss the technology, but to situate it accurately within the specific institutional, infrastructural, and regulatory landscape of the Indian public health system [3-6].

The Disease Burden and Why Conventional Surveillance Falls Short

India's vector-borne disease geography is complex and variable. Malaria transmission is concentrated in the high-burden states of Odisha, Jharkhand, Chhattisgarh, Madhya Pradesh, Maharashtra, West Bengal, and Assam, which together account for over ninety percent of national cases [7-10].

Dengue, by contrast, is now endemic in virtually every major Indian city and spreads into rural areas during monsoon season. Chikungunya and dengue often co-circulate in the same mosquito populations, making clinical differentiation difficult without laboratory testing. Japanese encephalitis has a strong geographic concentration in the Terai belt of Uttar Pradesh and Bihar, where waterlogged agricultural land provides extensive breeding habitat for Culex mosquitoes. Kala-azar, transmitted by the sandfly Phlebotomus argentipes, remains concentrated in the highly endemic districts of Bihar and Jharkhand, though elimination campaigns have reduced its geographic footprint considerably over the past decade [7-10].

The conventional surveillance response to this diversity relies on a hierarchy of passive reporting: peripheral health workers record cases using standardised forms, district health offices aggregate and transmit weekly data upward, and state health departments compile figures for the national programme. This system has several structural weaknesses that artificial intelligence cannot fix but can partially compensate for [7-10].

First, let us discuss the lag. A dengue outbreak in a Lucknow neighbourhood may produce a cluster of fever cases in the first week of August. Those patients attend private clinics, pharmacies, and government health centres in roughly equal proportions. The private sector cases are invisible to IDSP. The government health centre cases may be recorded on paper forms that reach the district office by the following Friday. The district compiles a weekly report that reaches the state the week after. By the time the signal is visible at state level, four weeks have passed. Four weeks is the difference between a nascent outbreak and an established epidemic [7-10].

Second issue is the denominator problem. India's officially reported dengue burden is a fraction of actual cases. Studies using sero-prevalence surveys and healthcare utilisation modelling consistently estimate that official figures represent between five and fifteen percent of true incidence. Models trained on reported case data therefore learn from a distorted picture of the epidemic, and their predictions inherit that distortion. A model trained on official dengue counts in a district where reporting captures one in ten cases will systematically underestimate transmission risk, and the bias will be worst in precisely the areas where healthcare access is poorest [7-10].

Third point is about geographic resolution. Disease patterns in Indian cities are highly localised. A neighbourhood with poor drainage, dense housing, and frequent water storage in open vessels can have dengue incidence ten times higher than a better-served neighbourhood two kilometres away. District-level surveillance averages this heterogeneity out of existence, making it impossible to target vector control precisely. A fogging truck dispatched to an entire district covers a lot of ground but may miss the three wards where the real problem is concentrated [7-10].

Fourth point is about the private sector gap. India's private healthcare sector handles a majority of outpatient care in urban areas and a substantial share in rural areas, but private providers are not systematically integrated into IDSP reporting. In states where private sector dengue treatment is concentrated in hospitals that charge fees the poorest patients cannot afford, the official data systematically under represents precisely the populations that face the highest mortality risk [7-10].

These are not minor technical complaints. They are the reasons why India's dengue caseload has increased every decade since systematic surveillance began, despite substantial investments in awareness campaigns and vector control. The surveillance system was not designed to generate the spatially and temporally granular data that effective outbreak response requires. Artificial intelligence can potentially help bridge that gap, though only if the proxy data sources it relies on are themselves adequate to the task [7-10].

AI Applications in the Indian Vector-Borne Disease Context

Outbreak Prediction and Forecasting

Machine learning-based dengue forecasting has been studied across several Indian cities including Delhi, Mumbai, Chennai, Pune, Kolkata, and Lucknow. The basic approach involves training regression or ensemble models on historical weekly case counts combined with meteorological variables, particularly rainfall, temperature, and relative humidity, which drive mosquito population dynamics and biting rates. The models learn the lagged relationship between climatic conditions and case counts, mosquito populations respond to rainfall a few weeks after a wet event, and human cases follow the mosquito population cycle by another fortnight or so, and then use current meteorological data to generate prospective forecasts [9-13].

Studies applying gradient boosting, random forest, and recurrent neural network architectures to Indian dengue data have found that models can generate weekly case count forecasts four to six weeks ahead with error margins that are operationally useful in a public health planning context. Six weeks of lead time is sufficient for a district health team to mobilise additional fogging equipment, intensify larval source reduction drives, pre-position stocks of dengue rapid diagnostic tests, and alert private hospitals to prepare for increased admissions. Whether any of these responses actually happen depends on administrative capacity and resource availability, not on the quality of the forecast, but the forecast is a necessary condition [9-13].

For malaria, the longest-running Indian forecasting work has been conducted in Odisha, the state with the country's historically highest malaria burden. Researchers have used recurrent neural network architectures trained on monthly rainfall totals, temperature data, and district-level case counts spanning multiple years to produce malaria incidence forecasts at four to eight week horizons. The Odisha work is notable not only for its methodological rigour but because it has begun moving from purely academic publication toward engagement with the state's National Vector Borne Disease Control Programme planning cycle. Translating a validated research model into a tool that district programme managers actually consult and act on is a separate and harder problem than building the model itself, and the Odisha experience is one of the few Indian examples where this translation has been seriously attempted, even if full operational deployment remains work in progress [9-13].

Japanese encephalitis forecasting in Uttar Pradesh has used similar approaches, with models incorporating paddy field extent derived from satellite imagery, temperature, and rainfall patterns in the Terai districts where transmission is concentrated. The disease's tight geographic concentration makes it somewhat easier to model than dengue, the relevant transmission zone is geographically circumscribed and the seasonal pattern is fairly predictable, and the catastrophic consequences of encephalitis outbreaks in the affected districts give these forecasts particular operational significance [9-13].

A methodological issue that runs through all of this forecasting work deserves explicit attention. Most published Indian dengue and malaria forecasting studies evaluate model performance on held-out historical data, the model is trained on years one through eight and tested on years nine and ten, or some equivalent split. This kind of retrospective evaluation gives an optimistic picture of real-world performance. In practice, a model deployed operationally faces data quality problems that do not appear in a carefully curated research dataset: delayed or missing case reports, broken weather station feeds, changes in diagnostic practice that shift what gets counted, and administrative boundary changes that disrupt time series continuity. The gap between retrospective accuracy and prospective operational reliability is real and has not been honestly reckoned with in the Indian AI health policy literature [9-13].

Satellite Remote Sensing and Risk Mapping

India has good satellite data coverage through the Indian Space Research Organisation's constellation of earth observation satellites, and ISRO has been an active participant in applying remote sensing to health-related applications. For vector-borne diseases, the most useful satellite-derived variables are land surface temperature, the Normalised Difference Vegetation Index, soil moisture, and surface water extent. Together, these proxies capture the environmental conditions that determine mosquito breeding habitat availability and population dynamics, standing water for egg-laying, vegetation that provides adult mosquito resting habitat, and thermal conditions that govern the rate of larval development and the speed of viral replication within the mosquito [10-16].

Studies applying satellite-derived habitat variables and machine learning to dengue risk mapping at sub-district level in Indian cities have demonstrated that these models can classify high-risk wards with accuracy substantially above baseline, and that the classifications have intuitive face validity. Wards with high proportions of construction sites, open drainage channels, and low-income housing without piped water supply consistently emerge as high-risk, consistent with what field entomologists would predict from first principles. The practical implication is that AI-driven risk maps could direct municipal fogging and larval source reduction operations to specific streets and neighbourhoods rather than treating entire districts uniformly, which is how most Indian municipal health departments currently operate [10-16].

In rural malaria-endemic districts, satellite-derived forest cover, irrigation canal networks, and paddy field extent have been incorporated into risk models to identify villages with the highest transmission potential. Forest cover is particularly important in the tribal belt of central India, where Plasmodium falciparum transmission is concentrated in communities living adjacent to forest margins and working in forest-based livelihoods. The National Remote Sensing Centre in Hyderabad has collaborated with the National Vector Borne Disease Control Programme on mapping exercises of this kind. The challenge is always the last mile: converting a risk map into an operational decision about which villages receive indoor residual spraying in a given month requires administrative systems and field capacities that the map alone cannot supply [10-16].

One methodological concern worth noting is that satellite-derived risk variables capture conditions at the time of image acquisition but may not reflect the highly dynamic nature of mosquito breeding site creation and elimination. A ward that is high-risk in satellite imagery taken in October may have changed substantially by November if drainage improvement works have been completed or if a construction site has been flooded. Static risk maps derived from infrequent satellite passes therefore overstate the precision of risk prediction in urban environments where the determinants of risk change quickly [10-16].

Social Media and Digital Surveillance

India's social media environment is large and linguistically diverse. As of January 2023, the country had approximately 467 million social media users, predominantly concentrated in urban and peri-urban populations. Several research groups have explored whether this data stream contains usable disease surveillance signals for dengue and other vector-borne conditions, motivated by the observation that people often discuss symptoms on social media before they seek formal medical care, and well before any case report reaches an epidemiologist [15-20].

Research applying natural language processing models to geo-tagged social media posts in Hindi and English from major Indian cities during monsoon seasons has found that symptom keywords and contextual language patterns can identify dengue-related discussion before official IDSP case counts begin to spike, in some analyses, by a margin of ten days or more. Ten days may sound modest, but in dengue epidemiology it is meaningful. The extrinsic incubation period of the virus in the mosquito is eight to twelve days, which means that a surveillance system detecting community-level transmission ten days before official confirmation is potentially catching signals close to the point of primary exposure. Intervention at that stage, rather than after the reporting pyramid has built up, is substantially more likely to contain an outbreak [15-20].

Google Trends data for dengue-related search queries has shown similar early-warning properties in the Indian context, with studies finding significant correlation between search volume for terms like 'dengue fever symptoms' and 'dengue treatment' and officially reported case counts at the state level. The correlation typically runs several weeks ahead of official notifications, suggesting that internet search behaviour reflects community-level awareness of dengue risk before the healthcare system has formally registered the outbreak [15-20].

The obvious limitation of social media and internet search surveillance is selection bias. These data streams reflect populations with smart phone access and digital literacy, which in India skews toward urban, younger, and higher-income groups. Rural populations in high-burden states like Odisha and Jharkhand, where the most severe malaria and kala-azar burdens are concentrated, are systematically underrepresented. A dengue early warning system that works well in Delhi and Mumbai but misses outbreaks in Bastar and Dumka is not an equitable solution to India's vector-borne disease problem, even if it is technically impressive within its coverage area [15-20].

There is also the misinformation problem. During the 2021 dengue surge in Delhi, social media carried a substantial volume of false information about papaya leaf extract as a dengue cure and about the risks of standard hospital treatment, alongside genuine symptom reports. Models that treat social media content as a proxy for disease incidence need to distinguish genuine symptom reports from misinformation amplification. Training classifiers to make this distinction requires labelled Indian-language datasets that are not yet publicly available at the scale needed for robust performance across Hindi, Bengali, Tamil, Telugu, and the other major languages in which disease-related discussion occurs online [15-20].

AI-Assisted Diagnosis

Diagnostic accuracy for vector-borne diseases at peripheral health facilities in India is poor. A primary health centre in a malaria-endemic district may have no functional laboratory, leaving the medical officer to make clinical diagnoses on the basis of symptoms that overlap substantially with other febrile illnesses. The classical dengue presentation, high fever, severe headache, retro-orbital pain, myalgia, and rash, is mimicked by chikungunya, leptospirosis, scrub typhus, and typhoid, all of which are endemic in many of the same areas. Rapid diagnostic tests for malaria are available but suffer from false negatives in low-parasitaemia infections, particularly for Plasmodium vivax. Dengue diagnosis beyond the NS1 antigen test requires ELISA or PCR, which are unavailable at most government health facilities below the district hospital level [20-26].

Deep learning models applied to blood smear microscopy have shown promising results in Indian settings. Con-volutional neural network architectures trained on large datasets of annotated blood smear images have demonstrated sensitivity and specificity for Plasmodium falciparum and Plasmodium vivax identification that approaches or matches the performance of expert microscopists, at least when the images are of adequate quality. In controlled pilots in malaria-endemic Indian districts, health workers with limited prior training have been shown to acquire images of adequate quality for automated analysis in the large majority of attempts, suggesting that image quality is not an insurmountable barrier to deployment at peripheral facilities [20-26].

For dengue, AI models that predict severe dengue risk from clinical parameters and complete blood count results have been developed and tested on datasets from Indian hospitals. The clinical utility of these tools is primarily in triage: a model that identifies which patients presenting with dengue fever are at high risk of progression to dengue haemorrhagic fever or dengue shock syndrome allows limited hospital beds and ICU capacity to be reserved for those who will most need them. In a district hospital with twelve ICU beds serving a catchment area of three hundred thousand people during a dengue surge, this kind of decision support has real value [20-26].

The deployment problem is real and underacknowledged. Most of these diagnostic tools were validated under research conditions in well-resourced hospital or research settings. The training data, whether blood smear images or clinical parameters, was collected under conditions of careful quality control that are not reproducible at the peripheral health post level. Performance in the field, on images taken by minimally trained workers using low-cost smartphones with variable lighting, or on clinical data entered by overworked nurses in a busy outpatient department, has not been systematically evaluated. The performance gap between validation accuracy and field reliability is a consistent finding in implementation science for digital health tools globally, and there is no reason to expect Indian AI diagnostics to be exempt from it [20-26].

Regulatory and Legal Framework

Data Protection under the DPDPA 2023

AI-based disease surveillance systems necessarily involve the collection and processing of personal health data. A patient whose fever consultation at a government health post is recorded in a digital system, combined with their location data and linked to other clinical records, has a personal data footprint that existing Indian law addresses imperfectly. The Digital Personal Data Protection Act 2023 is India's primary framework for data protection. It establishes consent as the principal basis for personal data processing, with narrow exceptions for state functions and legitimate purposes. Health data, including disease surveillance data, falls within the Act's scope. However, the Act's provisions for processing personal data for public health purposes are drafted broadly, and the subordinate regulations that would provide operational clarity have not yet been enacted as of the time of writing [21-30].

The Information Technology Act 2000 and the Sensitive Personal Data rules under it have provided some protection for health data in digital systems, but they were not designed with AI-driven population-level surveillance in mind. There are no requirements for algorithmic transparency, no obligations to disclose when AI predictions are being used to direct public health interventions, and no clear liability framework for cases where AI-driven surveillance fails to detect an outbreak that causes preventable deaths. The gap between what the law currently requires and what responsible deployment of an AI disease surveillance system would entail is significant [21-30].

The implications for data subjects are non-trivial. When a dengue risk mapping system uses anonymised geo-location data aggregated from mobile phones to identify high-risk neighbourhoods, the individuals contributing that location data have typically not consented to this use. When an AI malaria forecasting model incorporates individual patient records from government health facilities into its training dataset, the patients whose records are used have generally not been informed. These practices may be legal under a broad reading of state function exceptions, but they sit uneasily with the consent framework that the DPDPA 2023 establishes as the default. The absence of sector-specific guidance on health AI data processing is a gap that India's data protection authority will need to fill [21-30].

The Absence of AI-Specific Health Regulation

India does not yet have a regulatory framework specifically governing AI applications in healthcare or public health. The NITI Aayog's National Strategy for Artificial Intelligence published in 2018 identified health as a priority sector and called for responsible AI development, but it did not result in binding regulation. The Ministry of Health's National Digital Health Mission has focused on interoperability standards and infrastructure development rather than on the accountability questions raised by algorithmic decision-making in health. The regulatory space remains essentially empty [31-35].

The European Union's AI Act, which entered into force in August 2024, offers an instructive point of comparison. Under Article 6 of the Act, AI systems that are used as safety components of regulated medical devices, and that require third-party conformity assessment under EU medical device law, are classified as high-risk and subject to requirements including conformity assessments, technical documentation, data governance measures, and mandatory human oversight. It is important to note that the Act does not classify all health-related AI as high-risk: systems used for wellness monitoring, general health information, or administrative healthcare functions that do not qualify as medical devices are treated differently. India has not adopted analogous requirements. This creates a situation where AI tools for disease surveillance can be procured by state health departments and deployed at scale without any independent validation of their performance, any mandatory transparency about how their predictions are generated, or any ongoing monitoring for discriminatory or inequitable outputs [31-35].

The WHO's guidance on AI ethics in health, published in 2021, provides a normative baseline that includes principles of transparency, accountability, inclusion, and human autonomy in AI-assisted decisions. India's health authorities have referenced these principles in policy documents, but reference to principles is not the same as enforceable regulation. The accountability gap will become more visible as AI systems are deployed more widely in India's public health system, and closing it requires specific legislative or regulatory action rather than aspirational statements [31-35].

Liability Questions

The liability question in AI-assisted disease surveillance has not received adequate attention in Indian legal scholarship. Consider a realistic scenario: a state health department deploys a commercially procured AI early warning system for dengue. The system issues a forecast showing low risk in a particular district during August. Based on that forecast, the district health officer diverts fogging equipment and personnel to another district. An outbreak then occurs in the first district, causing avoidable deaths and morbidity. Who is legally responsible? [36-40].

Under Indian tort law, the analysis is unsettled. If the AI system was procured from a private vendor, product liability principles might apply, but Indian product liability law under the Consumer Protection Act 2019 was not drafted with algorithmic systems in mind and contains no provisions addressing the distinctive features of AI, adaptive models, probabilistic outputs, and the difficulty of attributing a specific output to a specific design choice. If the system was developed by a government agency, sovereign immunity considerations arise, although the doctrine's applicability to commercial-style government activities has been progressively narrowed by Indian courts. If the medical officer relied on the AI forecast but failed to apply independent clinical judgment, professional negligence principles may be relevant. None of these frameworks provides a clear answer to the scenario described [36-40].

The absence of clarity is not merely an academic concern. It creates perverse incentives at both ends of the adoption decision. Health administrators who cannot predict their legal exposure from AI reliance may be reluctant to adopt AI tools even when those tools would improve outcomes. Alternatively, and perhaps more likely, they may adopt them without establishing the accountability structures, independent validation, ongoing monitoring, clear protocols for when to override AI recommendations, that would catch systematic failures. Either response is worse than what a clear regulatory framework would produce. The law should clarify, before a major AI-assisted surveillance failure occurs, what standard of care applies to public health authorities using algorithmic decision support tools [36-40].

Infrastructure and Institutional Constraints

The most sophisticated dengue forecasting model in the world produces no public health benefit if the district health officer cannot access it, interpret its output, or act on its predictions in the time available. India's district health system has genuine capacity constraints that AI developers working primarily in university research environments have not always engaged with honestly. The translation from a validated model to an operationally deployed tool is not a matter of writing an API and waiting for adoption. It requires sustained engagement with the institutional, technical, and human systems that determine whether a prediction becomes a decision [38-43].

Internet connectivity at primary health centres and district health offices has improved substantially under the Digital India programme, but it remains intermittent in many rural and tribal areas. A dengue risk map that requires a cloud-connected browser to access is not useful to a medical officer whose connectivity drops for several hours each day. Models that require cloud connectivity for inference cannot run reliably in these settings. The offline-capable, compressed model variants that would work on a basic smartphone or a low-bandwidth connection have received considerably less research attention than the high-accuracy deep learning architectures that require server infrastructure. This is a choice with equity implications: the tools that work in urban referral hospitals get developed, and the tools that would work in remote primary health centres do not [38-43].

Health worker capacity is a related constraint. ASHA workers and auxiliary nurse-midwives, who are the front line of India's community public health system, typically have eight to ten years of formal schooling and limited training in digital tools beyond the specific applications they are required to use in their jobs. Introducing an AI-based malaria diagnostic tool into this setting requires not just software deployment but comprehensive implementation: training on image acquisition technique, on interpreting probabilistic outputs, on clinical decision protocols for acting on AI recommendations that differ from a worker's own assessment, and on what to do when the tool fails or produces an implausible result. The implementation science of deploying AI tools in community health worker settings in India has not been developed to match the pace of tool development which is a significant problem [38-43].

Data fragmentation is a structural problem that predates AI and will not be solved by it. India's health data sits in dozens of incompatible systems: IDSP for communicable disease surveillance, HMIS for health facility utilisation, IHIP for programme-specific disease reporting, state-level electronic health record systems in various states, and paper records at peripheral facilities that have not been digitised. An AI system that needs to integrate multiple data streams, dengue case reports, climate data, population data, healthcare utilisation patterns, to produce a useful forecast faces a data engineering problem that is often more challenging than the modelling problem itself. Data engineering is unglamorous, time-consuming, and poorly funded by the research grants that support model development. The result is that many Indian AI health models are built on cleaned, harmonised research datasets that do not exist in the operational environment where the model would need to be deployed [38-43].

The National Digital Health Mission aims to address data fragmentation through the Ayushman Bharat Digital Mission, which assigns unique health IDs and supports federated health records across providers. If this infrastructure develops as intended and achieves meaningful coverage, it will substantially improve the quality and accessibility of the data that AI disease surveillance systems need. But ABDM is at an early stage of deployment, its uptake has been uneven across states, and the timeline for achieving the coverage levels that would make federated AI modelling meaningful remains uncertain [38-43].

There is also an incentive problem within the research system. Indian academic researchers in AI and health are evaluated on publications in international journals, which rewards novelty in model architecture and strong validation metrics on clean datasets. Deployment-focused research, studying why a tool was not adopted, what training approach made field workers more likely to use an AI diagnostic correctly, or how a forecasting model's predictions should be communicated to a district health officer who has no statistical training, is less publishable and less rewarded. This creates a systematic bias toward the production of impressive research models and away from the patient institutional work of translating them into practice [38-43].

Algorithmic Bias and Equity Concerns

India's disease burden is not randomly distributed. Malaria disproportionately affects Scheduled Tribe populations in forested districts who face not only higher environmental exposure to transmission but also lower access to diagnosis and treatment. Dengue deaths are concentrated in populations without access to quality hospital care when they develop severe disease, the dengue fatality rate in well-resourced urban private hospitals is a fraction of the rate in under-staffed district hospitals and essentially zero in the tertiary facilities that serve the urban professional class. Kala-azar persists among the poorest rural households in Bihar and Jharkhand, where the socioeconomic conditions that sustain sandfly populations overlap precisely with the socioeconomic conditions that limit healthcare access. Any AI system that learns from historical data learns from a past in which these inequities already existed, and it will reproduce them unless its design actively and deliberately works against that tendency [35-43].

The risk of algorithmic bias in Indian health AI is underappreciated in the current policy discourse. Research from the United States has demonstrated convincingly that health risk prediction algorithms trained on healthcare utilisation data systematically underestimate risk in Black patients, because lower utilisation reflects barriers to access rather than lower health need, the algorithm interprets not-seeking-care as not-being-sick. Analogous patterns are entirely predictable in the Indian context. An AI dengue risk model trained on official IDSP case data will assign lower predicted risk to areas where under-reporting is high. Under-reporting correlates with healthcare access, which correlates with income. The model therefore systematically assigns lower predicted risk to low-income, low-healthcare-access areas, precisely the areas where true risk is likely highest. A public health system that acts on this model's outputs will allocate fogging resources and pre-positioned diagnostic supplies away from the communities most likely to be affected [35-43].

Designing against this bias requires deliberate choices at the data collection stage, the feature engineering stage, and the validation stage. Risk models should be validated specifically against seroprevalence surveys and community-based active surveillance studies, rather than only against official passive case counts, because official counts are not a neutral ground truth in the Indian epidemiological context. A model that performs well against official IDSP data but poorly against community-based seroprevalence data is not performing well, it is performing consistently with a biased training signal. This kind of equity-conscious validation is methodologically more demanding and more expensive than standard retrospective model evaluation, and it requires the kind of investment that most Indian AI health research projects, funded on short grant cycles, have not made [35-43].

There is also a language equity dimension that receives insufficient attention. Most AI disease surveillance tools developed for India operate in English or, at best, in Hindi. The populations that bear the highest vector-borne disease burdens speak Bengali, Odia, Gondi, Santali, and dozens of other languages in which symptoms are described differently and in which health-seeking behaviour is shaped by cultural frameworks that English-medium AI tools cannot engage with. A social media surveillance tool that monitors Twitter and Facebook in English and Hindi will miss the informal health communication networks in tribal and rural communities where disease burden is concentrated [35-43].

Policy Recommendations

The argument developed in this paper does not counsel against the use of AI in India's vector-borne disease management. The evidence for AI's technical potential is sufficiently strong, and the limitations of the existing surveillance system are sufficiently serious, that the question is not whether to pursue AI but how to pursue it responsibly. The following recommendations address the regulatory, institutional, and methodological gaps identified above [35-43].

First, the Ministry of Health and Family Welfare should develop an AI-specific addendum to the National Digital Health Mission framework that requires AI tools deployed in public health surveillance to disclose their training data sources, validation methodology, performance metrics, and known failure modes before procurement and deployment at scale. State health departments should be required to conduct a prospective performance assessment in a representative subset of districts before full deployment, with results published to a public registry. The conformity assessment approach in the EU AI Act's Article 6 offers a workable precedent, adapted to India's institutional realities [35-43].

Second, the Data Protection Board established under the DPDPA 2023 should issue sector-specific guidance on the processing of health data for AI-based disease surveillance. This guidance should clarify which data processing activities require explicit consent, which may proceed under state function exceptions, and what safeguards, including data minimisation, purpose limitation, and retention limits, apply specifically to AI-driven surveillance applications. The guidance should also establish a process for communities to seek information about how AI surveillance systems are being used in their areas and to raise concerns about discriminatory or inequitable outputs [35-43].

Third, the Indian Council of Medical Research should establish a public validation registry for AI diagnostic and predictive tools in vector-borne disease management, modelled on the Clinical Trials Registry India for pharmaceutical research. Tools seeking deployment in government health systems should be required to register their validation studies, including datasets, methodologies, and performance metrics disaggregated by geographic area and population subgroup, and to report ongoing performance data after deployment. This would create an evidence base that state health administrators could consult before procurement decisions, that researchers could use to identify gaps, and that civil society organisations could use to hold both vendors and health authorities accountable [35-43].

Fourth, medical education institutions should incorporate AI health tools into the training curricula for public health postgraduates, field epidemiology trainees, and district health managers. The relevant competencies are not technical, district health officers do not need to understand the mathematics of gradient boosting, but interpretive and critical: how to read a probabilistic forecast, what questions to ask about a model's validation, when an AI recommendation should be overridden by local knowledge, and what the liability implications of AI reliance are. These are competencies that the current generation of health administrators largely lacks, and that cannot be acquired without deliberate curriculum development [35-43].

Fifth, community health workers, ASHA workers, ANMs, and malaria technical supervisors, should be systematically involved in the co-design and field testing of AI tools intended for peripheral deployment. These workers know things about local data quality, patient behaviour, seasonal patterns, and operational constraints that no research team working on model architecture can replicate. Participatory design in this context is not only procedurally appropriate; it reliably produces tools that are better calibrated to the operational environment and that workers are more likely to use correctly and sustainably [35-43].

Sixth, research funding bodies, including the Indian Council of Medical Research, the Department of Biotechnology, and international funders operating in India, should explicitly prioritise implementation science for AI health tools alongside model development. A funded research agenda that asks why AI tools are not being adopted, how to train health workers to use them effectively, how to present probabilistic outputs to non-technical decision-makers, and how to detect and correct systematic bias in deployed models would address the translation gap that this paper has identified as the central obstacle to AI's contribution to India's vector-borne disease response [35-43].

CONCLUSION

India's vector-borne disease burden is large, historically persistent, and distributed in ways that track the deeper inequities of the country's social and economic geography. Artificial intelligence has something genuine to offer in addressing it. The evidence from Indian and comparable low-and-middle-income-country studies on dengue forecasting, malaria remote sensing, social media surveillance, and AI-assisted diagnostics is encouraging enough to justify serious investment, sustained institutional attention, and a willingness to work through the hard problems of data quality, implementation, and equity that stand between a validated research model and an operational tool that saves lives.

What the evidence does not support is the proposition that AI is a solution waiting only for adoption. The tools work best when the data systems that feed them are functional, when connectivity and hardware infrastructure are reliable, when health workers have adequate training and ongoing support, and when regulatory frameworks create accountability for failures and incentives for equity-conscious design. India has been building these foundations under the Digital India programme, the National Digital Health Mission, and successive iterations of the National Health Mission, but the foundations are not yet complete and their quality is uneven across the country's vast geographic and institutional diversity.

The risk in the current moment is that enthusiasm for AI in health among policymakers, technology companies, and international development agencies leads to large-scale deployment of inadequately validated tools in settings where they will be used by health workers without the training or support to interpret their outputs correctly, without regulatory frameworks to catch systematic failures, and without equity safeguards to prevent AI from making the existing mal distribution of healthcare worse. That path produces the worst of both worlds: the financial and reputational costs of AI adoption without the epidemiological benefits.

The better path requires honest validation that distinguishes retrospective accuracy from prospective operational reliability, equity-conscious design that validates against community-based surveillance rather than only official case counts, clear regulatory requirements that create accountability without extinguishing innovation, and patient integration with India's existing public health infrastructure rather than parallel AI-driven systems that bypass it. That path is slower and less photogenic than a press release about an AI-powered disease surveillance platform. It is also the one more likely to result in fewer dengue deaths.

Acknowledgement: The lead author thanks the co author for his contribution in the article.

Declaration: The lead author declares that the modalities given here is only suggestive in nature.

Funding: There was no funding received for the article.

Conflict of Interest: There is no conflict of interest regarding the article.

Foot Notes

1. Official NCVBDC data records approximately 193,245 confirmed dengue cases for the full year 2022. Some secondary sources citing lower mid-year figures reflect data as of partial-year cut-off dates rather than full-year totals.
2. NCVBDC official surveillance data records 83 confirmed malaria deaths in India in 2022, reflecting the well-documented limitations of passive case-finding and cause-of-death reporting. Modelled estimates from the Institute for Health Metrics and Evaluation's Global Burden of Disease Study place true malaria mortality for 2022 at approximately 9,400 deaths. The Medical Certification of Cause of Death (MCCD) system reported 1,265 malaria deaths for the same year. These differences reflect different methodologies and data sources and are consistent with the broader literature on disease burden estimation in low- and middle-income countries.
3. The estimate that India's official dengue surveillance captures between five and fifteen percent of true incidence is consistent with findings across multiple seroprevalence-based studies. For a systematic treatment of surveillance gaps and their implications for AI modelling which synthesises the broader literature on this point.
4. Research applying recurrent neural network and ensemble machine learning approaches to malaria incidence forecasting in Odisha using ICMR and state programme surveillance data has been carried out by multiple groups in collaboration with the Indian Council of Medical Research. For an overview of ICMR's activities in this area, see Indian Council of Medical Research,
5. Studies applying satellite-derived environmental variables and machine learning to sub-district dengue risk classification in Indian cities are reviewed in Nilavra Bhattacharya and others (n 4). For the underlying methodological approach to remote sensing-based dengue risk mapping, see also the environmental assessment work in central India reviewed in that systematic review.
6. Research applying NLP and social media data to disease surveillance in India has found lead times of approximately ten to fourteen days relative to official case notifications.
7. All of us should consult primary literature for city-specific performance parameters, as lead times vary by platform, disease, and reporting context.
8. As of January 2023, India had approximately 467 million social media user identities, equivalent to 32.8 percent of the total population.
9. One paper demonstrated that a widely used US health risk algorithm systematically underestimated need in Black patients by using healthcare costs as a proxy for health need, thereby treating lower utilisation, caused by access barriers, as lower clinical need.
10. For a representative overview of machine learning approaches to dengue forecasting using meteorological predictors and historical case data across Indian cities, see the primary studies surveyed in Bhattacharya and others (n 27). Performance metrics, including forecast horizon, error rates, and model architecture, vary substantially across individual studies and settings, and readers should consult primary literature for city-specific findings rather than treating any single study as representative.
11. The AI Act entered into force on 1 August 2024 and is applying progressively, with full obligations for high-risk AI systems scheduled for August 2026 (general) and August 2027 (medical devices). Article 6 defines high-risk AI as including systems that are safety components of products subject to third-party conformity assessment under EU harmonisation legislation, including medical devices regulated under Regulation (EU) 2017/745. Not all health-related AI systems are classified as high-risk under the Act.
12. A paper provides a systematic evaluation of convolutional neural network architectures for malaria blood smear image analysis. For Indian-context field deployment results, readers should consult implementation-focused primary literature, as the Xue paper is a methods evaluation rather than a field deployment study.
13. For AI-based triage and severity prediction models for dengue validated on Indian hospital data, see the literature surveyed in Bhattacharya and others (n 27). Individual hospital validation studies should be consulted for specific performance metrics, as results vary substantially by clinical setting, patient population, and model architecture.
14. Accredited Social Health Activist (ASHA), a health volunteer who is a married woman of the Local Self Governance (LSG), the selection criteria specify a minimum of Class 8 pass (approximately eight years of formal schooling), with Class 10 preferred in most states, resulting in a workforce with educational attainment typically in the range of eight to ten years of schooling.
15. For the applicable statutory framework, it is prudent to refer to the Consumer Protection Act 2019 of India which governs product liability but contains no AI-specific provisions; no settled judicial authority on AI product liability in the public health context exists under Indian law as of the time of writing.

REFERENCES

1. Park K, Text book of Preventive & Social Medicine, 27th edition, M. Banarasi Das Bhanot Publishers, Jabalpur, 2023, ISBN: 978-93-82219-19-4.

2. WHO, NTDS, https://www.who.int.
3. GOI, National Center for Vector Borne Diseases Control (NCVBDC), Dengue/DHF Situation in India, Dengue Cases and Deaths in the Country since 2017, Ministry of Health and Family Welfare, Government of India, 2023.
4. GOI, National Center for Vector Borne Diseases Control (NCVBDC), Magnitude of the Problem: Countrywide Epidemiological Situation 1995–2024, Ministry of Health and Family Welfare, Government of India.
5. World Health Organization, Vector-Borne Diseases, WHO Fact Sheet, 2 March 2020. https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases, accessed on 10 January 2024.
6. Bhattacharya N et al. Artificial Intelligence in Vector-Borne Disease Surveillance: A Systematic Review' of Neglected Tropical Diseases, PLOS, v17i4 e0010433.
7. GOI, NITI Aayog, Health System for a New India: Building Blocks, NITI Aayog, 2019.
8. GOI, ICMR, Annual Report, 2022-23.
9. Bhattacharya S et al. Use of Artificial Intelligence in Controlling Infectious Disease Outbreaks, Indian Journal of Community Medicine, v13, i02, pg112.
10. Chakraborty T et al. Now casting Disease Outbreaks Using Social Media Data, npj Digital Medicine, v6i54, 2021.
11. Das P et al. Tackling Dengue with Data: India's Early Warning System Pilot in Maharashtra, Journal of Health Informatics in Developing Countries, v10, i02, 2022.
12. Data Reportal, Digital 2023: India Data Reportal, January 2023. <https://datareportal.com/reports/digital-2023-india> accessed 15 January 2024.
13. GOI, Ministry of Health and Family Welfare, National Digital Health Mission, Blueprint, 2020.
14. Husnayain A et al. Applications of Google Search Trends for Risk Communication in Infectious Disease Management', International Journal of Infectious Diseases, v29, 2020.
15. Imran M et al. Using AI and Social Media Analytics to Monitor Disease Outbreaks in Low-Income Countries, Information Processing and Management, v129, 2020. pg102.
16. Rajib Paul et al. Remote Sensing and GIS in Epidemiology' in Ravi Mehrotra (ed), Geospatial Technology for Disease Mapping, Springer, 2019, 45.
17. Brady OJ et al. Refining the Global Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus, PLOS Neglected Tropical Diseases, v6, i08, 2012, e1760.
18. GOI, National Health Authority, Ayushman Bharat Digital Mission: Implementation Framework, 2022.
19. Reich NG et al. Challenges in Real-Time Prediction of Infectious Disease: A Case Study of Dengue in Thailand, PLOS Neglected Tropical Diseases, v11, i06, 2016, e0004761.
20. Rajput ZA et al. Evaluation of an Android-Based mHealth System for Population Surveillance in Developing Countries, Journal of Medical Internet Research, v14, 2012, e12.
21. GOI, Digital Personal Data Protection Act 2023, s 4. Ibid, s 7.
22. GOI, Information Technology Act 2000, s 43A, inserted by the Information Technology Amendment Act, 2008.
23. Obermeyer Z et al. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations, Science, v366, i 6464, 2019, pg 447.
24. Mittelstadt BD et al. The Ethics of Algorithms: Mapping the Debate, 2016.
25. GOI, NITI Aayog, National Strategy for Artificial Intelligence, 2018,
26. GOI, Ministry of Health and Family Welfare, Integrated Disease Surveillance Programme: Annual Report 2021-22.
27. Nilavra Bhattacharya et al. Artificial Intelligence in Vector-Borne Disease Surveillance: A Systematic Review, PLOS Neglected Tropical Diseases, v17, i04, 2022, e0010433.
28. GOI, Indian Council of Medical Research, Annual Report 2022-23.
29. World Health Organization, Ethics and Governance of Artificial Intelligence for Health: WHO Guidance, 2021.
30. Tshiab M et al. Artificial Intelligence and the Future of Global Health, The Lancet, 398, 1579, 2021.
31. EU, Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024. Artificial Intelligence Act, 2024, OJ L, 1689.
32. GOI, National Health Policy, para 4.2, 2017.
33. Xue H et al. Deep Learning for Image Recognition in Malaria Microscopy: Systematic Evaluation, PLOS Computational Biology, v14, i 08, e1008104.
34. Topol EJ, Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again Basic Books, 2019, pg 201.
35. Brownstein JS et al. Digital Disease Detection: Harnessing the Web for Public Health Surveillance, Science, 324, 2009, 989.
36. K Nsoesie et al. Forecasting Peaks of Seasonal Influenza Epidemics, PLOS Currents Outbreaks, v5, 2013.
37. Abiodun GJ et al. Modelling the Influence of Temperature and Rainfall on the Population Dynamics of Anopheles Arabiensis, Malaria Journal, v15, i 364, 2016.
38. Floridi L et al. An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles and

Recommendations, Minds and Machines, 28, 689, 2018.
39. GOI, Ministry of Health and Family Welfare, National Health Mission: Programme Implementation Plan 2021-22, pg31.
40. Zinszer K et al. Forecasting Dengue at the Neighbourhood Level: A Statistical Approach, Bulletin of the World Health Organization, v91, i 91, 2012.
41. GOI, Indian Council of Medical Research, Ethical Guidelines for Biomedical and Health Research Involving Human Participants,2017.
42. Lucey D, 'The Ethical Dimensions of AI in Global Health Surveillance, The American Journal of Bioethics, v20, i03, 2022.
43. Rahmatizadeh S et al. The Role of Artificial Intelligence in Management of Critical COVID-19 Patients', Journal of Infection, v101, 2020, pg 454.

Cite This Article: Navya Mall, Tridibesh Tripathy, Sanskriti Tripathy (2026). Role of Artificial Intelligence in Vector-Borne Disease Management in India. *EAS J Parasitol Infect Dis, 8*(2), 44-55.