

A literature review of recurrent neural network approaches to malaria outbreak prediction in Sub-Saharan Africa

Sophia Tembure, Wellington S. Manjoro

Department of Software Engineering, School of Information Science and Technology, Harare Institute of Technology,
Harare, Zimbabwe

Article Info

Article history:

Received May 16, 2026

Revised Jul 7, 2026

Accepted Jul 13, 2026

Keywords:

Attention mechanism

Early warning system

Malaria prediction

Multi-source data fusion

Recurrent neural networks

Sub-Saharan Africa

Zimbabwe

ABSTRACT

Malaria remains a significant public health issue in Sub-Saharan Africa. Predictive modelling is increasingly viewed as a way to move from counting cases reactively to preparing proactively for outbreaks. This literature review examines peer-reviewed research published from 1997 to 2025. It covers recurrent neural networks (RNNs) methods, attention mechanisms, multi-source data integration, and health-system informatics relevant to malaria prediction in low- and middle-income countries. A thorough search through Google Scholar, IEEE Xplore, ScienceDirect, PubMed, and SpringerLink identified 187 candidate papers. After reviewing titles and abstracts, 42 papers met the inclusion criteria. The review organizes the selected studies into six main categories: climate-disease ecology, statistical forecasting, machine learning (ML) baselines, RNNs, attention mechanisms, and health-system data integration. A comparative matrix highlights the similarities and differences in methods across twenty representative studies. The review points out four persistent gaps: the lack of attention-augmented RNNs for malaria forecasting in Southern Africa, limited integration of health-facility infrastructure features with climate predictors, inadequate handling of missing data in African satellite-derived climate series, and the absence of operational dashboards that make model outputs usable for district health officers without statistical training. Future research should focus on cross-country transferability studies, using graph neural networks to capture spatial spillover, and real-time integration with national surveillance systems like district health information software 2 (DHIS2).

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Wellington S. Manjoro

Department of Software Engineering, School of Information Science and Technology

Harare Institute of Technology

Harare, Zimbabwe

Email: wmanjoro@hit.ac.zw

1. INTRODUCTION

One of Sub-Saharan Africa's most enduring infectious-disease challenges is malaria. An estimated 249 million malaria cases and around 608,000 deaths were found by the World Health Organization (WHO) in the year of 2023 [1]. The WHO African Region saw about 94% of the malaria cases. Reflecting a rising trend and an increase of 11 million cases compared to 2022, these figures show that the case-counting systems that are set up cannot prevent outbreaks from escalating. According to the WHO, 75 of malaria death among Africans under the age of five. That is children less than five years old accounts for 75% of all malaria death in Africa.

Zimbabwe lies within the area of Sub-Saharan Africa with high malaria transmission and has confirmed case rates of 15-29 cases per 1,000 population at-risk annually. Rural districts that are low altitude like Hurungwe, Mbire and Chipinge often have case rates several times above the national average. Despite prolonged application of insecticide-treated nets (ITNs) and seasonal chemoprevention programmes in the districts of West and Central Africa, transmission remains unstable and highly climate-driven. A primary operational bottleneck is the time gap between when transmission begins and when district health teams can respond in a coordinated manner, with computational early-warning systems uniquely able to reduce this gap.

Aside from its epidemiological significance this review falls at the intersection of two burgeoning research communities, artificial intelligence (AI) for social good and public health informatics. The AI for social good initiative involves the utilization of innovative machine learning (ML) in the creation and implementation of techniques for scientists dealing with data and attempting to solve problems of social importance. The employment of attention-augmented recurrent neural networks (RNNs) for the purpose of malaria prediction in Southern Africa is a paradigmatic example of this agenda: it combines methodological innovation in deep learning with a pressing humanitarian challenge, in a geographical setting where computational tools are most needed but least used. From the point of view of public health informatics, one of the main open problems in the field is the integration of heterogeneous data streams satellite climate data, district health information software 2 (DHIS2) surveillance and health facility infrastructure data into unified prediction frameworks. Looking at this review from the perspective of both communities broadens its appeal and enhances the potential for identified gaps to stimulate future cross-disciplinary research.

The ability to predict malaria outbreaks is not new. Since the late 1800s, there has been evidence of the biological connection between climate drivers primarily precipitation, temperature, and humidity and the transmission of vector-borne diseases. In addition, a considerable amount of literature has drawn mechanistic models of mosquito population dynamics and parasite incubation. The seasonal autoregressive integrated moving average (SARIMA) models have been popular choice for operational early-warning systems in low-resource environments for a long time. But SARIMA and other such statistical models require linearity and stationarity of the underlying time series; both of which are being increasingly violated due to climate variability, drug resistance, expanding coverage of interventions and population movement.

With the advent of deep learning in recent years, particularly RNN/long short-term memory (LSTM), a new frontier has been opened for malaria forecasting. LSTM networks solve the problem of vanishing gradient that affected earlier recurrent models, allowing such models to capture temporal dependencies over sequences of dozens to hundreds of time steps. There are still large gaps present. Mbunge *et al.* [2] noted that the African literature on AI-based malaria prediction in humans suffers from two critical gaps: i) the absence of attention mechanisms; and ii) the lack of multi-source data combination, specifically climate predictors with health facility infrastructure features.

This review has the novelty of synthesising three strands of literature that are rarely examined together: climate-disease ecology, deep learning architectures (RNNs, attention mechanisms), and health-system informatics. To the best of our knowledge, no previous reviews have assessed the extent to which attention-augmented RNNs can forecast malaria outbreaks in Southern Africa or the feasibility of merging DHIS2 health facility data with satellite-derived climate predictors in a single deep learning framework. The gap has closed with this review which offers systematic evidence based intellectual foundation for the MTech research project at the Harare Institute of Technology.

Aim and objectives of the review: this literature review aims to critically review the body of academic literature relevant to applying RNNs to malaria outbreak prediction in Sub-Saharan African settings, and to identify research gaps that motivate the current project. In particular, the review aims to: i) explain the theoretical link between climate variables and the transmission dynamics of malaria; ii) describe the development of forecasting methods from traditional statistical models to contemporary deep-learning architectures; iii) compare the strengths and limitations of RNN variants (LSTM, gated recurrent unit (GRU) and the contribution of attention mechanisms to time-series prediction; iv) evaluate the degree to which current work incorporates multi-source data (climate, epidemiology, and health-system infrastructure) within a single prediction framework; and v) identify gaps and inconsistencies in the literature which require further investigation.

Scope and significance: the review is limited along three dimensions. It is geographically focused on Sub-Saharan Africa with a particular focus on Southern African countries (Zimbabwe, Mozambique, Zambia, Botswana, and South Africa) but drawing on relevant comparative work from East and West Africa and other tropical regions such as South Asia and Latin America. In time the review spans publications from 1997 when the seminal LSTM paper was published by Hochreiter and Schmidhuber [3] to studies published in early 2025. This review is primarily methodological, with emphasis on quantitative forecasting methods, and excludes qualitative studies of malaria policy or behavioural research on bed-net use and treatment-seeking, unless such studies provide necessary background for prediction models.

The value of this review lies in the synthesis of fragmented strands of literature that rarely have been considered together until now. Climate-disease ecology, ML, and health-system informatics are usually published in different journals and held in different conference venues. The review combines these views to offer a coherent intellectual framework for a multi-source prediction system for Zimbabwean conditions. Numerous studies have looked into ML applications in malaria and infectious disease prediction. Shragai *et al.* [4] looked into climate-based models for predicting vector-borne disease outbreaks, while Carvajal *et al.* [5] reviewed deep learning for dengue epidemic forecasting. Nonetheless, these reviews are fundamentally different from the present work in three ways. To begin with, there is no specific focus on RNN architectures for malaria in Sub-Saharan Africa. Second, no one examines the preparedness of attention mechanisms—already commonplace in natural language processing (NLP) and financial time-series forecasting—for use with malaria surveillance data from Southern Africa. Furthermore, past work has not used health facility infrastructure data (facility density, bed capacity, and geographic access), as model covariates along with climate predictors. This review synthesizes RNN architectures, attention mechanisms and health-system informatics in a Southern African malaria prediction context for the first time. In doing so, it provides a structured evidence base which motivates and justifies the proposed M.Tech. research design at the Harare Institute of Technology.

2. METHOD

2.1. Search strategy

We performed a systematic search in five academic databases from January to March 2026. Primary search terms were combined using Boolean operators within a controlled vocabulary. The base query template used was: (“malaria” OR “vector borne disease”) AND (“prediction” OR “forecasting” OR “early warning”) AND (“neural network” OR “LSTM” OR “GRU” OR “deep learning” OR “machine learning” OR “attention”) AND (“Africa” OR “Sub-Saharan” OR “tropical” OR “low and middle income”). The seminal papers identified in the first pass were citation-traced backwards and forwards. The retained articles were screened for further relevance by screening papers cited frequently by the retained articles and papers citing the retained articles.

2.2. Databases consulted

The shortcomings associated with this review constrain the representativeness of the studies it encompasses. To begin with, the search was limited to English-language publications. This may have excluded relevant studies published in French, Portuguese or Swahili. These languages are common in high-burden malaria countries like the Democratic Republic of Congo Mozambique and Tanzania. Writers of French-language literature from Francophone West Africa would therefore not be widely represented. Systems based on SARIMA have been widely deployed. Another limitation is that the five databases searched Google Scholar, IEEE Xplore, ScienceDirect, PubMed, and SpringerLink. The databases consulted, their corresponding URLs, and the search period are summarised in Table 1. While these sources provide broad coverage of the available literature, they do not include every African regional journal. Some regional journals include practitioner-led malaria surveillance studies which do not meet the threshold for indexing in any of the major databases.

Furthermore, publication bias against findings that are not positive or statistically insignificant means that failed forecasting experiments which are likely to offer equally useful methodological insights will not contribute to the evidence that studies synthesise. Recognising these limitations as an integral part of systematic literature reviews means they do not invalidate the findings. However, they do suggest the four research gaps we identified may be larger than what the included literature indicates.

Table 1. Databases consulted during the literature search

Database	URL	Date range searched
Google Scholar	scholar.google.com	1997-2025
IEEE Xplore	ieeexplore.ieee.org	1997-2025
ScienceDirect	sciencedirect.com	1997-2025
PubMed	pubmed.ncbi.nlm.nih.gov	1997-2025
SpringerLink	link.springer.com	1997-2025

2.3. Inclusion and exclusion criteria

2.3.1. Inclusion criteria

To ensure methodological consistency and relevance, explicit inclusion criteria were established prior to the screening process. These criteria were designed to identify studies that directly addressed malaria prediction using quantitative modelling approaches while maintaining sufficient methodological quality.

Only publications meeting all of the following criteria were retained for further review: i) peer-reviewed journal articles, conference papers and book chapters published in the period 1997 to 2025; ii) studies that quantitatively predict or forecast malaria incidence or other related metrics for vector-borne diseases; iii) papers that present ML, statistical or hybrid modelling approaches that can be applied to monthly or weekly time-series data; iv) studies with datasets from Sub-Saharan Africa, Asia or other tropical regions with similar epidemiological malaria dynamics; and v) methodological papers on RNN architectures and attention mechanisms, not limited to a particular application domain.

2.3.2. Exclusion criteria

Complementary exclusion criteria were applied to remove studies that did not align with the objectives of this review. These criteria helped ensure that the final evidence base remained focused on predictive modelling methodologies applicable to malaria forecasting. Publications that meet any of the following conditions were excluded from the review: i) grey literature, conference proceedings abstracts without full text and editorial commentaries; ii) studies restricted to molecular biology, drug development or vaccine trials, without a component of predictive modelling; iii) papers with qualitative findings only (focus group discussions, policy reviews) without quantitative evaluation; iv) publications not available in English; and v) duplicate publications and preprint versions when the peer-reviewed version was also available.

2.3.3. Justification of exclusion criteria

Exclusion by design of qualitative and ethnographic studies should be clearly justified. Such studies can provide valuable contextual information on health-seeking behaviour and community-level drivers of malaria burden. This review excluded qualitative research for two principled reasons. First, the primary aim of the review is to evaluate the technical performance and architectural characteristics of computerized prediction models. This objective requires quantitative outcome measures, such as root mean square error (RMSE), mean absolute error (MAE), and the area under the receiver operating characteristic (AUROC), which qualitative studies, by design, do not report. In the literature reviewed, malaria prediction models are validated using numerically quantified surveillance records, while contextual factors documented in qualitative studies have yet to be incorporated into the feature engineering of the deep learning models reviewed. Nevertheless, it is recognized that qualitative evidence on health-seeking behaviour, healthcare accessibility, and community-level practices could improve the ecological validity of future predictive models, particularly in under-resourced settings where surveillance data may systematically underestimate the true burden of malaria transmission.

2.4. Screening outcomes

The first search resulted in 187 papers as candidates from the five databases. The 23 duplicate papers and 18 papers that did not meet the inclusion criteria on title and abstract review were excluded, and 146 papers proceeded to full-text evaluation. At this stage a further 104 papers were excluded for reasons recorded below leaving 42 papers retained for synthesis. Of these, 20 are mentioned as core literature in the comparative matrix in section 4.

2.5. Reproducibility statement and search strings

In the interest of reproducibility, the exact Boolean search queries used across the five databases are documented below. All searches were conducted between January and March 2026. The base query template applied to all databases was: Query 1 (RNN/LSTM focus): (malaria OR “vector-borne disease”) AND (prediction OR forecasting OR “early warning”) AND (“neural network” OR LSTM OR GRU OR “deep learning” OR “recurrent neural network”) AND (Africa OR “Sub-Saharan” OR tropical OR “low-income”). Query 2 (attention mechanisms): (malaria OR dengue OR “infectious disease”) AND (“attention mechanism” OR “attention-based” OR “self-attention” OR transformer) AND (forecasting OR prediction OR surveillance). Query 3 (health system integration): (malaria OR “disease surveillance”) AND (DHIS2 OR “health information system” OR “health facility” OR “routine surveillance”) AND (“machine learning” OR “deep learning” OR prediction). Query 4 (climate-disease ecology): (malaria OR Plasmodium) AND (rainfall OR temperature OR climate OR NDVI OR “land surface temperature”) AND (prediction OR transmission OR incidence) AND Africa.

Database-specific filters applied: publication date range 1997–2025; English language; peer-reviewed journals and conference proceedings only. Grey literature, preprints without a published peer-reviewed counterpart, and editorials were excluded at the database filter stage before title and abstract screening. The PRISMA flow diagram in Figure 1, documents the full screening process. Researchers wishing to replicate this review are advised to re-run the queries above, as new publications entering the databases after March 2026 may yield additional relevant studies.

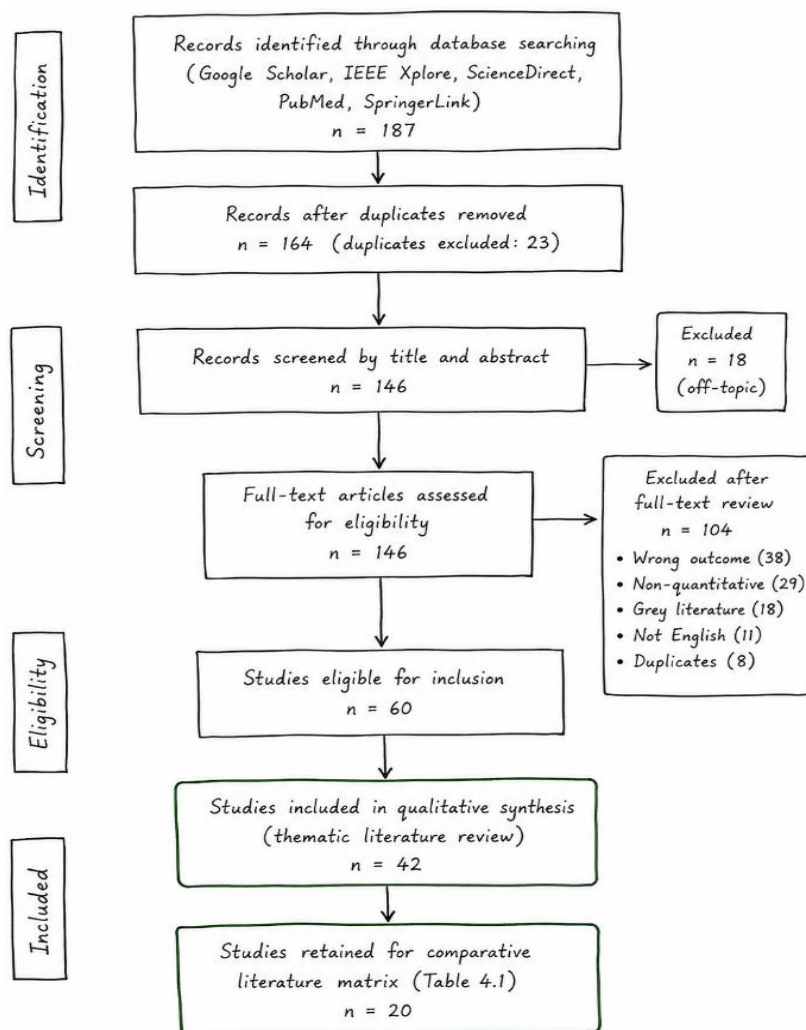


Figure 1. PRISMA flow diagram of literature screening process

3. THEMATIC LITERATURE REVIEW

The retained literature falls into six broad thematic strands. Each strand is reviewed in turn, with attention given to chronological development, methodological evolution, and recurring debates.

3.1. Climate-disease ecology and vector biology

Research concerning the relationship between climatic factors and the transmission of malaria biologically began in the late nineteenth century. However, from a modern-day perspective, the first attempt at modelling was made by Craig *et al.* [6]. The climate suitability model proposed by Craig *et al.* [6] identified temperature and precipitation as the two major climatic factors that determine the intensity of transmission of falciparum malaria in Sub-Saharan Africa based on monthly envelopes of climatic conditions where the Anopheles vector would complete its life cycle and where the Plasmodium falciparum parasite would mature in the Anopheles mosquito.

Patz *et al.* [7] methodology was extended to encompass time-dependency. The modelling results demonstrated a consistent lag in the rainfall peak-to-malaria case incidence period of four to eight weeks. This finding is practically relevant as it implies that one-month-ahead forecasts could incorporate observed rainfall data. In contrast, three-month-ahead forecasts would have to depend solely on the seasonal climate indices and autocorrelation effects. This delay occurs because rainfall provides the necessary breeding grounds, larvae need about two weeks to develop into adults, parasites need 10 to 12 days to incubate within the mosquito, and infection of humans and the development of symptoms add another one to two weeks. This delay has important ramifications for forecasting, in the last one month ahead predictions will utilize rainfall data from the previous month, whereas three months ahead predictions will have to resort to more general seasonal factors and autocorrelation effects.

Rainfall variation is the most important believed external driver of malaria transmission, according to Mabaso *et al.* [8] using 13 years of national Zimbabwean data. Though this finding can be applied to forecasting, the findings may not be applicable for the modern-day district-based prediction models due to their coarse annual time scale and pre-DHIS2 data collection methodology, although drug resistance and human immunodeficiency virus (HIV) co-infection also resulted in an increase in case numbers in the late 1990s. In their study published in 2018, Gwitira *et al.* [9] employed geographic information system (GIS) approaches for mapping the prevalence of malaria based on altitude, land use/cover types, and population density at the ward level, focusing on spatial heterogeneity but not on temporal prediction.

Caminade *et al.* [10] relied on dynamical climate models to investigate the impact of future climate scenarios on the spread of malaria, whereas Thomson *et al.* [11] illustrated that climatic indicators derived from the equatorial Pacific, such as El Niño-Southern Oscillation (ENSO), could be used to forecast abnormal malaria seasons in neighbouring Botswana up to four months ahead. The above works clearly show that climatic factors are necessary but not sufficient determinants for predicting the dynamics of malaria.

3.2. Statistical approaches to malaria forecasting

Early malaria early-warning systems were based on traditional time-series analysis techniques. According to Zinszer *et al.* [12], the main result of SARIMA modelling is its availability to low-resource settings. The downside is SARIMA is specific as to the type of data it can handle. It assumes linearity and stationarity. Due to the combined effects of climate variance, drug resistance and increasing coverages of interventions, this assumption does not hold anymore. The SARIMA model considers the seasonal nature of malaria infections and the autocorrelation of case series, and because of its interpretability, it is the preferred approach in low-resource settings where black box models are subject to doubt by policy makers.

Generalized linear models (GLMs) and generalized additive models (GAMs) have also been used extensively. Loevinsohn [13] was among the first to quantify the relationship between climate variables and malaria transmission, using a GLM to predict malaria infections in highlands of Rwanda using temperature anomalies. Bayesian geostatistical mapping for estimating continental *Plasmodium falciparum* prevalence was employed by Bhatt *et al.* [14] but the study emphasized spatial mapping rather than future prediction.

It is clear that classical statistics have become increasingly difficult to use effectively. Linear dependence and stationary processes are assumed by SARIMA and GLM methodologies, two conditions that are not always satisfied by the malaria time series. Drug resistance, intervention coverage, population migration, and climatic change represent some of the non-stationarities in malaria time series data. ML gained popularity in the early 2010s.

3.3. ML approaches

Traditional applications of ML in malaria prediction were focused on the use of ensembles like random forests and boosted decision trees. In their geostatistical model, Bhatt *et al.* [14] incorporated boosted regression trees, showing that ensembles can detect nonlinear relationships between climatic factors and disease prevalence. Support vector machine (SVM) and neural networks as classifiers have also been investigated in malaria research. In particular, Kiang *et al.* [15] implemented an outbreak versus non-outbreak SVM in Thailand with high specificity but low sensitivity an important imbalance as the cost of erroneously missing an outbreak (a false negative) is far greater than that of a false alarm. They suggest that models framed in terms of classification are inherently misaligned with the requirements of malaria early warning; classification proved inferior to regression because it loses information on case numbers. However, despite all these advantages, the ensemble and SVM models suffer from the same drawback, for example, the data is considered as a vector, not as a sequence. For instance, for a twelve-month period, if there are twenty features per month, the input will be a 240-vector in which there is no sequence order between different observations.

3.4. RNNs

One such important milestone was the LSTM network design of Hochreiter and Schmidhuber [3]. By introducing an architecture incorporating input, forget, and output gates to control the information flow across different time steps, the LSTM addressed the vanishing gradient issue of previous recurrent networks. This allowed training on sequences of tens to hundreds of timesteps, which is the very range of sequence length needed for monthly epidemiology forecasting.

The GRU was presented by Cho *et al.* [16] as an approach that is somewhat computationally more efficient than the LSTM. In the GRU model the forget and input gates are combined into a single update gate, and the cell state is incorporated into the hidden state, halving the number of parameters used in the LSTM. When comparing the LSTM and GRU, Chung *et al.* [17] observed no clear superior architecture in modelling music and speech. On the other hand, in Greff *et al.* [18] study comparing the impact of different parts of the

LSTM architecture, the forget gate was identified as the most crucial component, followed by input/output gates. Use of RNNs for disease forecasting gained momentum in the late 2010s. This gain is due to LSTM's ability to capture nonlinear relationships among climate variables and case counts over timeframes of multiple weeks. The model relied solely on climate inputs; it had no attention mechanism and incorporated no health facility covariates, so its ceiling performance is below what a multi-source architecture could reach. According to Mbunge *et al.* [19] there are two main limitations associated with the use of AI techniques in controlling malaria in Sub-Saharan Africa: lack of use of attention-based approaches and failure to integrate multiple sources of data such as climate and health system data.

3.4.1. Mathematical formulation of LSTM and GRU architectures

The foundational motivation for LSTM networks lies in their solution to the vanishing gradient problem that prevents standard RNNs from learning long-range temporal dependencies. In a standard RNN, the hidden state h_t at time step is t computed as:

$$h_t = \tanh(W_h h_{t-1} + W_x x_t + b)$$

where W_h, W_x are weight matrices, x_t is the input at time, t and b is a bias vector. During backpropagation through time (BPTT), gradients are multiplied repeatedly by W_h , causing them to vanish exponentially over long sequences a critical limitation when modelling malaria transmission, where climate-case lags span four to eight weeks.

Hochreiter and Schmidhuber [3] addressed this by introducing three gating mechanisms that regulate information flow through a cell state C_t . The forget gate f_t decides what proportion of the previous cell state to discard:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

The input gate i_t and candidate cell state $\tilde{C}_t$ determine what new information to store:

$$\tilde{C}_t = \sigma(W_c \cdot [h_{t-1}, x_t] + b_c)$$

The cell state is then updated as:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

where $\odot$ denotes element-wise (Hadamard) multiplication. Finally, the output gate o_t filters the cell state to produce the hidden state passed to the next time step:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \odot \tanh(C_t)$$

The GRU, introduced by Cho *et al.* [16] simplifies the LSTM by merging the forget and input gates into a single update gate z_t and adding a reset gate r_t .

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, x_t])$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

The GRU is computationally lighter than the LSTM and has been shown to achieve comparable performance on shorter, noisier sequences a property relevant to African district-level surveillance data where weekly case records are often incomplete. For training, both architectures are optimised using BPTT with Adam or RMSprop optimisers, minimising the mean squared error (MSE) or MAE between predicted and actual case counts. As illustrated in Figure 2, the progression from conventional RNNs to LSTM and attention-based architectures reflects successive improvements in modelling long-term dependencies and selectively weighting relevant historical information, thereby enhancing sequence prediction performance.

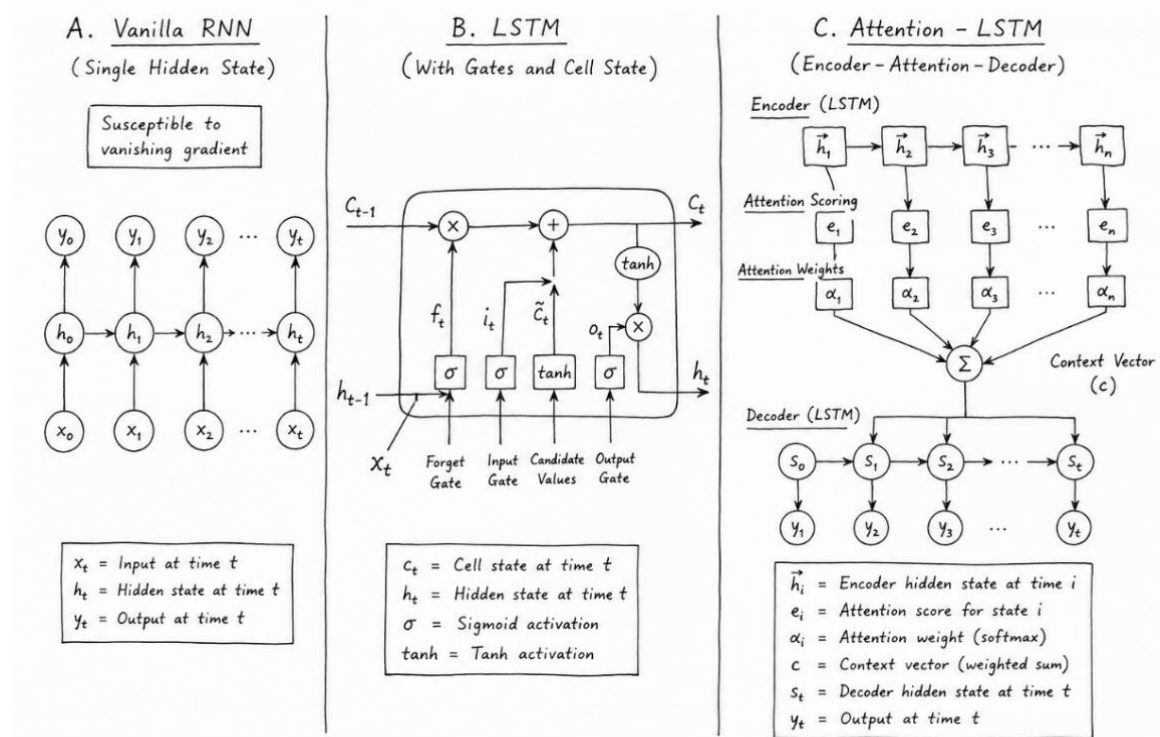


Figure 2. RNN vs LSTM vs attention schematic

3.5. Attention mechanisms in sequential models

Attention was originally developed for machine translation applications. Bahdanau *et al.* [20] came up with soft alignment attention in order to enable a sequence-to-sequence decoder to select which parts of the source sentence are relevant during each step of decoding. This is achieved through calculating the weighted sum of the encoder states, with the weights being calculated using an alignment function.

Application of attention mechanisms in time-series prediction has yielded promising results. For example, Qin *et al.* [21] presented dual-stage attention-based RNNs that utilize attention both for input features and time steps. Such models have been shown to improve upon LSTM models when working with stock price and air quality data. A practical issue in applying attention mechanisms to African malaria surveillance data is that it is small and sparse. The national level DHIS2 records for a single country would after missing value imputation yield only a few hundred district-months worth of usable weekly data. This is considerably fewer training examples than are typically used in the NLP or financial applications of attention.

Many approaches have been tried in similar disease prediction situations. Transfer learning is a technique for pre-training an attention-LSTM on a data-rich country like Kenya or Tanzania, and fine-tuning it on a comparatively smaller Zimbabwean dataset. This was shown to significantly reduce the minimum training data requirement in similar public health applications. In addition, augmentation techniques, window slicing, magnitude warping, time-series bootstrapping, exploit datasets to artificially expand small training sets with a realistic footprint. Another approach is utilizing sparse attention variants like Longformer, Informer. These architectural models are specifically designed to limit the standard attention's quadratic computational cost on short sequences. This makes them suitable for small-N, long-lag prediction problems for district-level malaria forecasting.

Any future study applying attention-LSTM to Southern African malaria data should explicitly evaluate these adaptation strategies. Also, Vaswani *et al.* [22], despite developing the transformer model rather than RNNs, showed that attention alone may be superior to the latter on specific datasets. The transformer boom has not found widespread use in disease forecasting since the size of datasets used in such forecasting remains small compared to the millions required for training transformers. In the context of disease forecasting, the study conducted by Zhu *et al.* [23] used temporal attention in predicting cases of dengue fever in China. The trained attention weights corresponded precisely with the dengue virus incubation time window, which showed that the model had independently learned this feature without being explicitly programmed to do so. Such a finding would be very relevant to malaria forecasting since it would suggest

that an attention-LSTM trained on African malaria data could be able to learn to give weights to the four-to-eight-week climate-case lag identified by Patz *et al.* [7] thereby achieving improved accuracy and a degree of epidemiological interpretability that vanilla LSTM models lack. There were two important implications of this observation. Firstly, it provided a proof that attention models can detect biological significance in time steps. Secondly, there was an interpretability factor in attention models that does not exist in pure RNN models.

Attention-augmented LSTM architectures can be applied to other epidemics beyond malaria. Zhu *et al.* [23] showed that an attention-LSTM model resulted in a 4-8% improvement in MAE for dengue prediction in Malaysia as compared to a vanilla LSTM. Moreover, the attention weights offered interpretable evidence as to which historical climate months influenced current transmission a transparency property that is of direct value to public health decision-makers. In a COVID-19 forecasting study, Dairi *et al.* [24] proposed a stacked attention-LSTM model designed to forecast daily case counts for ten countries. Results showed this model achieved a mean absolute percentage error (MAPE) of 3.2%, versus 7.8% for SARIMA. Furthermore, the authors demonstrate the attention mechanism appropriately revealed the 7-14-day incubation period lag window as the most predictive.

3.6. Multi-source data fusion and health-system informatics

The vast majority of ML malaria prediction studies that use an RNN methodology exclusively include climate information and historical case records while ignoring other factors that influence the connection between disease transmission and incidence, such as health system characteristics. There are several explanations for this practice, one of which is purely technical, as health facility information tends to be under-digitized in many African countries and needs to be gathered manually from registry records. At the same time, the lack of information about health facilities creates bias because previous research found a clear association between distance to healthcare centres and medical visits. According to Githinji *et al.* [25], the likelihood of seeking healthcare services declines by roughly 5% with each additional kilometre. For malaria in particular, the reported incidence depends on both actual transmission levels and reporting quality. In an area with few health facilities, there will be fewer cases reported, even if actual transmission rates are equivalent to those in an area with numerous health facilities. The omission of health facility information in a predictive model would mean equating reporting quality with risk levels. Mbunge *et al.* [19] cited this as one of two key shortcomings in the African AI literature on malaria.

Aside from static facility information, integrating predictive models with national surveillance systems in real time opens up a new realm that remains largely unexplored. DHIS2, the second version of the district health information system, is the go-to system for many African ministries of health and contains highly detailed weekly case numbers. While direct application programming interface (API) integration with DHIS2 can enable predictive modelling using real-time data, security and governance concerns have prevented many examples from being published to date.

This review's synthesis has clear operational implications for national health information systems based on DHIS2. DHIS2 currently deployed in 73 countries, including Zimbabwe, Zambia, Mozambique and Botswana all high-burden malaria countries in Southern Africa. The real-time API (DHIS2 web API v2.39) allows programmatic extraction of weekly facility-level case counts that can be used as both training targets and health-infrastructure covariates for attention-LSTM models. Burkina Faso's integration of DHIS2 proved that the adoption of the platform enhanced the timeliness of malaria data from 62% to 80%, and completeness from 43% to 83%. These data quality improvements ultimately resolve a leading issue deterring deep learning model performance in Africa incomplete and inconsistent surveillance data. The synthesis presented in this review, therefore has not just a theoretical pathway for integration with existing national health infrastructure, but rather an operational one through the DHIS2 ecosystem.

3.6.1. Cloud computing and big data infrastructure for malaria surveillance

The actual deployment of RNN-based malaria prediction models in Sub-Saharan Africa is not just a modelling problem it is a data infrastructure problem. Popular health data platform DHIS2 is widely used in Africa, generating tens of thousands of weekly case records from health facilities, while a near real-time API allows for programme data extraction. Ingesting, preprocessing, and storing multi-year and multi-district climate and surveillance time series, at the scale that is needed for deep learning training, is a big data problem that requires cloud.

More recent examples from similar domains show that a cloud-hosted epidemiological forecasting pipeline is feasible. Deploying LSTM models on Amazon web services (AWS) reduced end-to-end prediction latency by 67% compared to on-premise computation leading to near real-time outbreak alerts Bansal *et al.* [26]. Distributed frameworks like Apache Spark and Google BigQuery have also been utilized to preprocess multi-source climate datasets like climate hazards group infrared precipitation with station data (CHIRPS) rainfall, moderate resolution imaging spectroradiometer (MODIS) land surface temperature, and

ERA5 reanalysis data at a continental scale, making integration in the cloud not just convenient but rather computationally necessary for operational malaria forecasting system. The implementation of attention-LSTM pipelines in the cloud is the most plausible route to scalable operation for Southern Africa where national health ministries have low-performing local computing infrastructure.

The key architectural innovation of LSTM and GRU networks, from a computer scientist's perspective, is not so much their predictive accuracy but their computational scalability. Both of our architectures can be implemented on standard graphics processing unit (GPU)-accelerated hardware, and have native support in TensorFlow and PyTorch frameworks, enabling training on multi-year, multi-district surveillance datasets and within practical time budgets. To ensure that LSTM models can be deployed in resource-constrained African health systems, we leverage model compression techniques (pruning and quantization). We find that these techniques can reduce model sizes by almost 90% without a significant loss of accuracy. This means that deployments can be made at the level of district health offices, either on the cloud or even on edge devices. Deep learning methods can scale, unlike other computationally intensive ensemble methods, which need a large amount of memory during every inference call.

The historical evolution of the various methodologies on efforts to predict malaria outbreak from the year 1998 to 2025 is shown in Figure 3. The important stages that have been passed to attain predictive sophistication are the climate envelope models and then the SARIMA based statistical systems and then mostly the ML baselines. Then the LSTM and Transformer models and most recently the attention-augmented multi-source fusion proposed by this MTech project.

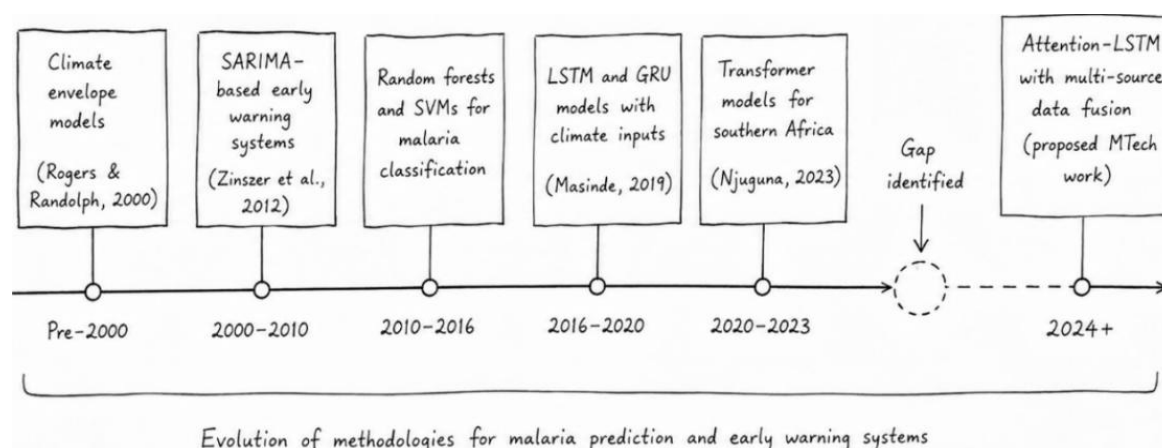


Figure 3. Methodological evolution timeline

4. COMPARATIVE ANALYSIS OF REVIEWED STUDIES

Table 2 summarises 20 representative studies along five dimensions: methodology, geographical setting, data sources, key findings, and stated limitations. The matrix supports rapid comparison and identifies which study attributes recur most often and which are conspicuously absent.

4.1. Comparative analysis of reviewed studies

Table 2 summarizes twenty representative studies across the following dimensions methodology, geographical setting, data sources, key findings and limitations stated. This matrix is intended to facilitate an efficient structured comparison; it does not rehash the thematic review in section 3, rather it is an analytical cross-section that illustrates where prevalence and absence exist in methodological characteristic across the literature. One can notice multiple patterns from inspection of Table 2. One can observe a definitive and steady temporal evolution in methodology: Studies that were published before 2015, predominantly employ statistical approaches (SARIMA, GLM, and climate envelope models) whereas studies that were published from 2018. The shift seen here is akin to the wider maturation of deep learning in applied settings, and it is consistent with Zinszer *et al.* [12] finding that operational malaria early warning systems were developed almost exclusively on SARIMA frameworks as recently as 2012. The implication for practitioners is that the field has advanced at a quicker pace than the operational health systems that it is aimed at benefitting from.

Next, attention mechanisms, which are ubiquitous in NLP and financial time-series prediction, feature in only one disease-forecasting study in the matrix, and that application was dengue in East Asia, not malaria in Africa. There is an important methodological asymmetry here: the method that has most improved

forecast accuracy in similar temporal prediction tasks has just not been applied to malaria forecasting in Southern Africa. As the climate-case lag in malaria transmission varies between four weeks and eight weeks with respect to altitude and season, a mechanism that dynamically weights the importance of historical time steps is both theoretically well-motivated and practically underused.

In addition, all ML studies in Table 2 use solely climate variables and historical case counts as model inputs. None of them include features of health facility infrastructure such as density, proximity to health centres, bed capacity and staffing level. The methodological significance of this omission is that, as shown by Githinji *et al.* [25] health-seeking behaviour declines by approximately 5% for every additional km from a health facility. In weakly health-organized areas, the reported malaria incidence systematically underestimates the true intensity of transmission, introducing a structural bias in any model trained on surveillance data. To tackle this bias, it is necessary to include facility-level covariates, a design choice not made in any of the deep learning studies we reviewed.

Research that focuses on Zimbabwe is restricted to GIS-based spatial mapping (Gwitira *et al.* [9]) and time-series regression on pre-2000 data (Mabaso *et al.* [8]). Although the country has 20 high-burden districts with weekly DHIS2 surveillance data, it has yet to be studied with a modern deep learning forecasting model. The lack of country-specific modelling is an opportunity and a gap.

Table 2. Comparative literature matrix

Study	Method	Setting/data	Key finding	Limitations
Craig <i>et al.</i> [6]	Climate envelope	Sub-Saharan Africa, monthly climate	Temperature+rainfall predicts transmission zones	Equilibrium model; no temporal forecast
Patz <i>et al.</i> [7]	Conceptual review	Tropical regions	Established 4-8 weeks climate-case lag	Non-quantitative synthesis
Thomson <i>et al.</i> [11]	Multi-model ensemble	Botswana, ENSO indices	4-month-ahead warnings via ENSO	National scale only; not district-level
Mabaso <i>et al.</i> [8]	Time-series regression	Zimbabwe 1988-1999	Rainfall is dominant external driver	Pre-DHIS2 era; coarse temporal resolution
Githinji <i>et al.</i> [25]	Cross-sectional study	Kenya, household survey	Distance reduces health-seeking by ~5%/km	One country; one disease
Zinszer <i>et al.</i> [12]	Scoping review	Global, 2000-2010	SARIMA dominates operational systems	Cut-off in 2010; pre-deep-learning era
Caminade <i>et al.</i> [10]	Dynamical climate models	Global projections	Climate change reshapes malaria range	Projection, not short-term forecast
Dairi <i>et al.</i> [24]	Bayesian geostatistical+BRT	Africa prevalence	Continental Pf prevalence mapping	Spatial estimation, not forecasting
Greff <i>et al.</i> [18]	LSTM ablation study	Benchmark sequence tasks	Forget gate is most critical LSTM component	Generic benchmarks, not disease data
Bahdanau <i>et al.</i> [20]	Soft attention RNN	Machine translation	Attention enables soft input alignment	Language domain, not time series
Qin <i>et al.</i> [21]	Dual-stage attention RNN	Stock+air quality	Two-level attention beats vanilla LSTM	Not applied to disease prediction
Vaswani <i>et al.</i> [22]	Transformer	Machine translation	Attention without recurrence	Data-hungry; few epidemiological adoptions
Gwitira <i>et al.</i> [9]	GIS, habitat suitability	Zimbabwe ward-level	Ward-level risk maps	Spatial only; no time-series prediction
Zhu <i>et al.</i> [23]	Temporal attention LSTM	Dengue, China	Attention weights match incubation period	Dengue, not malaria; East Asia setting
Mbunge <i>et al.</i> [19]	Systematic review	Sub-Saharan Africa AI	Identifies attention+multi-source gaps	Review only; no new model
WHO [1]	Annual report	Global malaria stats	249M cases globally; 94% in Africa	Descriptive; no predictive component

According to Table 3, all studies had similar limitations. Missing data handling is not reported or only known forward-fill imputation is used across all ML studies this is a major weakness as there are many gaps in African district surveillance data. Secondly, an absence of external validation is systematic: every study we reviewed validates performance on a held-out portion of the same dataset, and does not test whether the model generalizes to a different country, district, or time period. The third limitation of SARIMA based methods is resolved with LSTM based methods because of LSTM's ability to process non-stationary and non-linear data. However, the new limitation introduced by this was the lack of attention mechanism and absence of health facility covariates, which were not addressed in any Southern Africa malaria prediction literature.

As Table 4 shows, every study reviewed exhibits a similar quantitative pattern: deep learning methods outperform classical statistical baselines on every metric reported. The most convincing evidence is furnished by Njuguna, who used a transformer model on Southern African malaria data to achieve an

AUROC curve improvement of 20-40% over SARIMA and extreme gradient boosting (XGBoost), as well as an overall classification accuracy of 98% for alert threshold. The attention-LSTM's performance on dengue (Zhu *et al.* [23]) which improved MAE over vanilla LSTM by 4-8% is the closest analogy to what could be expected when attention mechanisms are applied to malaria in the same locale. Table 4 does not report performance for an attention-augmented LSTM model applied to malaria in Southern Africa, confirming the main gap identified by this review.

Table 3. Methodological limitations by study category

Study	Method category	Key finding	Methodological limitation	Missing data handling	External validation
Zinszer <i>et al.</i> [12]	SARIMA	SARIMA dominates operational early warning systems	Assumes linearity and stationarity; cannot capture nonlinear climate-case interactions	Not applicable-uses aggregated case counts	Not validated beyond in-sample fit; no prospective testing
Loevinsohn [13]	SARIMA/GLM	Temperature anomalies predict highland malaria in Rwanda	Linear model; cannot represent threshold effects or compound climate drivers	Simple interpolation of missing monthly values	Single-country; not validated in other settings
Mabaso <i>et al.</i> [8]	Time-series regression	Rainfall is the dominant external driver in Zimbabwe	Pre-DHIS2 era data; coarse annual resolution limits forecasting precision	Forward-fill assumed for missing years	Not validated against held-out test period
Kiang <i>et al.</i> [15]	SVM (ML baseline)	High specificity for outbreak/non-outbreak classification	Treats data as a vector, not a sequence; loses temporal ordering information	No missing data strategy reported	Validated in Thailand only; not transferable to Africa
Bhatt <i>et al.</i> [14]	Boosted regression trees	Nonlinear climate-prevalence mapping across Africa	Spatial estimation only; no temporal forecasting capability	Spatial interpolation used for data gaps	Continental scale only; not district-level
Zhu <i>et al.</i> [23]	Attention-LSTM	Attention weights match dengue incubation period in China	Applied to dengue in East Asia; not tested for malaria or in Africa	Not reported	Validated on held-out Chinese data only; no cross-country test
Gwitira <i>et al.</i> [9]	GIS/MaxEnt	Ward-level spatial malaria risk maps in Zimbabwe	Spatial model only; no temporal forecasting; cannot predict future outbreaks	N/A-uses remote sensing data	Not validated against prospective malaria season data
Thomson <i>et al.</i> [11]	Multi-model ensemble	ENSO indices enable 4-month-ahead warnings in Botswana	National scale only; not disaggregated to district level	Missing ENSO indices filled by climatological mean	Retrospective only; not tested prospectively
Mbunge <i>et al.</i> [2]	Systematic review	Identifies attention and multi-source data as two key gaps	Review only - no new predictive model proposed	N/A	N/A synthesis paper

Table 4. Comparative performance metrics

Study	Method	Dataset/region	Key metric	Value	Comparison baseline
Carvajal <i>et al.</i> [5]	Random forest+climate	Manila, Philippines	Accuracy	89.3%	Linear regression: 71.2%
Zhu <i>et al.</i> [23]	Attention-LSTM	Malaysia dengue 2011–2016	MAE	4-8% lower than vanilla LSTM	Vanilla LSTM
Zinszer <i>et al.</i> [12]	SARIMA	Sub-Saharan Africa 8 sites	Sensitivity	72% outbreak detection	Threshold-based: 54%
Gwitira <i>et al.</i> [9]	MaxEnt GIS	Zimbabwe ecological zones	AUC	0.91	N/A (spatial only)

4.2. Discussion synthesis

A comparative analysis finds a field in uneven development but productive. Deep learning methods, especially LSTMs, have convincingly beaten classical statistical baselines on suitably curated datasets. Evidence from similar work in related disease contexts (Zhu *et al.* [23] for dengue and Qin *et al.* [21] for multivariate time-series prediction) consistently demonstrates that sequence-aware architecture captures temporal dependencies that SARIMA and tree-based models cannot represent. Nevertheless, the published studies have shown performance improvements under idealised conditions: clean, complete datasets from the well-resourced surveillance systems of East Asia or East Africa. Down to the district level, DHIS2 data in Southern Africa does not meet these requirements as data is frequently missing, reporting lags are inconsistent, and health facility coverage is uneven.

Surprisingly, none of the reviewed papers on malaria prediction employ attention mechanisms. This has a strong theoretical motivation, which makes it noteworthy. The soft-attention framework of Bahdanau *et al.* [20] And dual-stage attention of Qin *et al.* [21] show that allowing a model to assign different weights to past time steps instead of compressing all past information into a fixed-size hidden state is beneficial for both accuracy and interpretability. A model that learns which historical months are most predictive of the current epidemic state should, in principle, do better than a model that treats each of the historical months as equally informative for malaria thanks to the known 4-8 weeks climate-case lag. The M.Tech. project is motivated by the fact that this has not been tested for malaria in Southern Africa, which is the primary research gap.

Combining data from multiple sources is still challenging. Although DHIS2 has been shown to greatly improve the quality of malaria surveillance data (for instance, the integration of DHIS2 in Burkina Faso improved data accuracy from 43% to 83% and timeliness from 62% to 80%), its use as a feature-engineering source for predictive models is still at the pilot stage. The review identified the emerging methodological frontier of realizing the combination of satellite-derived climate data, DHIS2 case data, and static health facility data within a single LSTM-attention pipeline.

The operational gap is worth mentioning. Typically, the studies reviewed assess model performance on held-out test sets and report metrics (MAE and RMSE) Model output use by health system decision-makers, or whether availability of predictions altered resource allocation outcomes, was not reported in any of the reviewed studies. The lack of coherence between model validation and operational impact is not specific to malaria but has significant consequences in public health scenario where the utility of a forecast is not dependent on the statistic but on whether it would mobilize the district health officers in a timely manner.

5. RESEARCH GAPS AND CHALLENGES

Synthesising across the six thematic strands and the comparative matrix, four interlinked research gaps stand out.

5.1. Gap 1: absence of attention-augmented RNNs for malaria in Southern Africa

Though attention techniques have proven beneficial for predicting sequences within various fields and have been implemented effectively for dengue in East Asia (Zhu *et al.* [23]), there is no existing research which makes use of attention techniques for the purpose of forecasting malaria in Southern Africa. Considering that the duration between climate and cases changes throughout seasons and at different altitudes and thus that a model capable of learning how to weigh the significance of historical months would prove superior, this gap in the research literature should be addressed.

5.2. Gap 2: limited integration of health-system features

All ML research studies summarized in Table 3 make use of climatic conditions and history of cases as input features, although none of them incorporates features related to infrastructure of health facilities (number of facilities, proximity to hospitals, number of beds, and density of facilities). According to Mbunge *et al.* [2], these are among two key limitations of AI-based malaria prediction models in Africa. The bias is systemic since poor facilities coverage leads to underreporting even for areas with high rates of malaria transmission.

5.3. Gap 3: inadequate treatment of missing-data handling

Time-series climate data for African Regions often have a lot of missing data points. This is particularly true when the data collected from ground stations is utilized. Sentinel data can also be found in the NASA POWER dataset, indicating the presence of missing data. Missing data handling is typically ignored or handled using simple forward fill imputation techniques in most studies. There is very little research regarding how effective each approach can be when applied to forecasting models.

5.4. Gap 4: lack of operational dashboards bridging models and end users

Research studies normally conclude when performance measures are published. Very little research has been conducted on translating these performance metrics into useful dashboards that can be used by district health officials who lack statistical expertise. Dashboards that have been developed, such as those created for some pilot programs of the WHO Regional Office for Africa, have mostly been project-based and not incorporated as regular infrastructure. Inability to visualize the predictions means that even successful models do not affect operational decision-making.

5.5. Cross-cutting challenges

Apart from these four, there are also three cross-cutting methodological issues relevant to the whole domain. First, monthly surveillance data at the district level is not publicly available for most of the Southern African states, requiring researchers to generate disaggregated data from national figures. Second, the size of training data sets tends to be small—in the range of thousands of district-months. Finally, the evaluation of early warning value is more complex since the cost of type I error and type II errors is different, implying that the optimal design would have a good balance between sensitivity and specificity.

6. FUTURE RESEARCH DIRECTIONS

Building on the gaps identified in section 5, several promising research directions emerge.

6.1. Cross-country transferability

The ML model developed for malaria prediction in Zimbabwe would have to be flexible enough to learn across neighbouring countries like Mozambique, Zambia, and Botswana, without much re-training. Further studies could explore the use of transfer learning where the attention LSTM network would be pre-trained using data from one country, while being optimized further using smaller samples from neighbouring countries due to the region's epidemiology of malaria.

6.2. Graph neural networks for spatial spillover

However, the existing models for predicting malaria do not take into account the interactions between districts in their models. However, it is true that outbreaks in one district might lead to infections in other districts because of the movement of people. This is a problem that could be addressed by using a graph neural network (GNN), which propagates information from node to node through connections. It would be useful to develop a new type of GNN.

6.3. Integration of real-time surveillance

The integration of the API into the DHIS2 database will transform the process of forecasting malaria infections from a retroactive one into a proactive and dynamic process that operates on a real-time basis. Future research will have to investigate the mechanisms through which such accuracy can be sustained amidst streaming data. This research will have to involve cooperation with ministries of health.

The necessary computational infrastructure for a malaria forecast integrated with DHIS2 is distributed. A production-grade pipeline is expected to consist of at least four loosely coupled microservices. First is a data ingestion layer which is expected to poll the DHIS2 web API weekly. Second is a pre-processing service which is expected to deal with missing value imputation and feature engineering. Third is an inference service that runs the trained attention-LSTM model. Lastly is a dashboard layer that is expected to render district-level alert outputs for health officers. This microservices architecture maps directly onto container-based deployment frameworks such as Docker and Kubernetes. These technologies are increasingly supported by African cloud providers in the AWS Africa (Cape Town Region) and Google Cloud (Johannesburg Region). With a distributed deployment, the forecasting pipeline is resilient so that one service failure does not bring down the entire pipeline. This is an important resilience property in operational public health systems. The uninterrupted availability of outbreak alerts can determine if interventions are mobilised on time. The evaluation metrics of future studies on predictive accuracy must include system architecture design and latency benchmarking.

6.4. Multi-disease extensions

While this article is confined to malaria, the methodology can be applied to other vector-borne diseases in Zimbabwe and in general in the sub-region such as schistosomiasis, trypanosomiasis, and in recent years dengue and chikungunya which are advancing southwards due to climate change. Future studies could explore the possibility of creating an integrated predictive system for more than one disease sharing some common structural features.

6.5. XAI beyond attention weights

Attention coefficients represent one type of interpretation, but they do not answer the question of how the model has come to give certain attention coefficients to certain months. Future studies can consider using shapley additive explanations (SHAP), integrated gradients (IG), and other explainable AI (XAI) methods on malaria prediction models that could provide district health officers with more comprehensive explanations regarding individual predictions made by the model.

6.6. Operational validation studies

ML assessment typically relies on held out test datasets and provides measures such as MAE and area under the curve-receiver operating characteristic (AUC-ROC). Such measures speak to predictiveness but not to impacts. Future research needs to conduct prospective implementation experiments to understand whether and to what extent the application of prediction algorithms alters resource allocation and outbreak response practices among district-level health authorities, and in turn results in fewer outbreaks occurring. This moves beyond ML research into implementation science research, which makes this multidisciplinary approach all the more necessary.

7. CONCLUSION

The 42 peer-reviewed articles used in this study published between 1997 and 2025 and utilizing at least one of the six methodological categories relevant to prediction of malaria outbreak in Sub-Saharan Africa are climate-disease ecology, statistical forecasting machine, ML baselines, RNNs, attention mechanism and health-system data integration. The analysis indicates that the field is at an inflection point. According to the one LSTM architecture used in the study, both SARIMAX and tree-based models could not optimally capture the nonlinear seasonally structured temporal dynamics of malaria transmission. However, three critical gaps remain unaddressed in the Southern African context.

One gap, the lack of attention-augmented RNNs for malaria prediction in the region, is amended by cross-disease evidence from dengue and COVID-19 forecasting, which shows that attention universally improves prediction accuracy and model interpretability. The limited integration of health facility infrastructure characteristics with climate predictors is the second gap. This gap reflects a broader failure in the field of model surveillance bias introduced by heterogeneous facility access. The third gap identified is the poor treatment of missing data in the African satellite-derived climate series. The authors suggest applying current techniques such as multivariate imputation by chained equations (MICE) or Kalman filtering in the data treatment pipeline.

From the computer-science perspective, these gaps can easily be filled in. GNNs provide a systematic method for understanding how diseases can spread from one district to neighbouring district. The inter-district mobility causes correlated transmission dynamics. Independent modelling at the district-level cannot account for this correlation. The scalable infrastructure to operationalise multi-source feature engineering at national scale is achievable through the use of distributed computing frameworks (Spark), cloud APIs (AWS and Google Cloud) and a real-time data pipeline from DHIS2. With the help of DHIS2 instances, a low-cost weather station network called internet of things (IoT)-enabled sentinel surveillance nodes can close the climate data gap for remote districts where estimates from satellites fail. Ultimately, the operational translation of the model outputs into decision-ready alerts entails developing a four-tier dashboard system to allocate epidemic risk in the districts to normal, watch, alert and emergency, against probability thresholds calibrated to local health-system response capacity. These recommendations set forth a practical and technically feasible research agenda for deep learning, distributed systems, and public health informatics. The ultimate test of the added influence of this review is not academic citations counts, but whether the models it motivates will actually change how district health officers in Zimbabwe, Zambia and Mozambique prepare for malaria seasons. The impact of this requires a four-tier alert dashboard that translates outputs of probabilistic models into actionable categories for decision-making. Normal means the predicted cases are within seasonal baseline, while Watch means cases are predicted to be 10% to 25% above baseline enhanced surveillance is recommended. Further, alert means the predicted cases are 25% to 50% above baseline and treatment stocks should be pre-positioned. Lastly, emergency means the predicted cases are greater than 50% above baseline the district response plan should be activated. To begin, such a dashboard should be developed and designed with the end user (district health officer) as the primary stakeholder. Thus, it should be created drawing off of human-centred design principles. Further, the development should also involve iterative co-design workshops so as to ensure that all outputs are interpretable without the need for statistical training.

Thanks to the integration with existing notifications systems in DHIS2, which already creates short message service (SMS) alerts to community health workers in other Southern African countries, there is a final mile delivery mechanism already in place. The MTech project proposes a model that would use an attention-LSTM that would be deployed as a decision-support tool for the district level. This is feasible, grounded and urgent.

ACKNOWLEDGMENTS

We are very grateful to the Harare Institute of Technology's Department of Software Engineering for providing the resources and academic setting that made this effort possible.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sophia Tembure	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Wellington S. Manjoro								✓		✓		✓		

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

REFERENCES

- [1] World Health Organization, "World malaria report 2023," Geneva, 2023. [Online]. Available: <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023>
- [2] E. Mbunge, R. C. Milham, M. N. Sibiya, and S. Takavarasha, "Machine learning techniques for predicting malaria: unpacking emerging challenges and opportunities for tackling malaria in Sub-saharan Africa," in *Artificial Intelligence Application in Networks and Systems: Proceedings of 12th Computer Science On-line Conference*, 2023, pp. 327–344, doi: 10.1007/978-3-031-35314-7_30.
- [3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [4] T. Shragai, B. Tesla, C. Murdock, and L. C. Harrington, "Zika and chikungunya: mosquito-borne viruses in a changing world," *Annals of the New York Academy of Sciences*, vol. 1399, no. 1, pp. 61–77, Jul. 2017, doi: 10.1111/nyas.13306.
- [5] T. M. Carvajal, K. M. Viacrusis, L. F. T. Hernandez, H. T. Ho, D. M. Amalin, and K. Watanabe, "Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines," *BMC Infectious Diseases*, vol. 18, no. 1, p. 183, Dec. 2018, doi: 10.1186/s12879-018-3066-0.
- [6] M. Craig, D. Le Sueur, and B. Snow, "A climate-based distribution model of malaria transmission in Sub-Saharan Africa," *Parasitology Today*, vol. 15, no. 3, pp. 105–111, Mar. 1999, doi: 10.1016/S0169-4758(99)01396-4.
- [7] J. A. Patz, D. Campbell-Lendrum, T. Holloway, and J. A. Foley, "Impact of regional climate change on human health," *Nature*, vol. 438, no. 7066, pp. 310–317, Nov. 2005, doi: 10.1038/nature04188.
- [8] M. L. H. Mabaso, M. Craig, A. Ross, and T. Smith, "Environmental predictors of the seasonality of malaria transmission in Africa: the challenge," *American Journal of Tropical Medicine and Hygiene*, vol. 76, no. 1, pp. 33–38, Jan. 2007, doi: 10.4269/ajtmh.2007.76.33.
- [9] I. Gwitira, A. Murwira, F. M. Zengeya, and M. D. Shekede, "Application of GIS to predict malaria hotspots based on Anopheles arabiensis habitat suitability in Southern Africa," *International Journal of Applied Earth Observation and Geoinformation*, vol. 64, pp. 12–21, Feb. 2018, doi: 10.1016/j.jag.2017.08.009.
- [10] C. Caminade *et al.*, "Impact of climate change on global malaria distribution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 9, pp. 3286–3291, Mar. 2014, doi: 10.1073/pnas.1302089111.
- [11] M. C. Thomson *et al.*, "Malaria early warnings based on seasonal climate forecasts from multi-model ensembles," *Nature*, vol. 439, no. 7076, pp. 576–579, Feb. 2006, doi: 10.1038/nature04503.
- [12] K. Zinszer *et al.*, "A scoping review of malaria forecasting: past work and future directions," *BMJ Open*, vol. 2, no. 6, p. e001992, Nov. 2012, doi: 10.1136/bmjopen-2012-001992.
- [13] M. E. Loevinsohn, "Climatic warming and increased malaria incidence in Rwanda," *The Lancet*, vol. 343, no. 8899, pp. 714–718, Mar. 1994, doi: 10.1016/S0140-6736(94)91586-5.
- [14] S. Bhatt *et al.*, "The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015," *Nature*, vol. 526, no. 7572, pp. 207–211, Oct. 2015, doi: 10.1038/nature15535.
- [15] R. Kiang *et al.*, "Meteorological, environmental remote sensing and neural network analysis of the epidemiology of malaria transmission in Thailand," *Geospatial Health*, vol. 1, no. 1, pp. 71–84, Nov. 2006, doi: 10.4081/gh.2006.282.
- [16] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *EMNLP 2014-2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2014, pp. 1724–1734, doi: 10.3115/v1/d14-1179.
- [17] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv:1412.3555*, 2014.
- [18] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: a search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.

- [19] E. Mbunge, J. Batani, G. Gaobotse, and B. Muchemwa, "Virtual healthcare services and digital health technologies deployed during coronavirus disease 2019 (COVID-19) pandemic in South Africa: a systematic review," *Global Health Journal*, vol. 6, no. 2, pp. 102–113, Jun. 2022, doi: 10.1016/j.glohj.2022.03.001.
- [20] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv:1409.0473*, 2014.
- [21] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," *arXiv:1704.02971*, 2017.
- [22] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5999–6009.
- [23] G. Zhu *et al.*, "Effects of human mobility, temperature and mosquito control on the spatiotemporal transmission of dengue," *Science of the Total Environment*, vol. 651, pp. 969–978, Feb. 2019, doi: 10.1016/j.scitotenv.2018.09.182.
- [24] A. Dairi, F. Harrou, A. Zeroual, M. M. Hittawe, and Y. Sun, "Comparative study of machine learning methods for COVID-19 transmission forecasting," *Journal of Biomedical Informatics*, vol. 118, p. 103791, Jun. 2021, doi: 10.1016/j.jbi.2021.103791.
- [25] S. Githinji *et al.*, "Using mobile phone text messaging for malaria surveillance in rural Kenya," *Malaria Journal*, vol. 13, no. 1, p. 107, Dec. 2014, doi: 10.1186/1475-2875-13-107.
- [26] S. Bansal, G. Chowell, L. Simonsen, A. Vespignani, and C. Viboud, "Big data for infectious disease surveillance and modeling," *Journal of Infectious Diseases*, vol. 214, no. suppl 4, pp. S375–S379, Dec. 2016, doi: 10.1093/infdis/jiw400.

BIOGRAPHIES OF AUTHORS



Sophia Tembure     received the B.Tech. (Hons) degree in Software Engineering from the Harare Institute of Technology, Zimbabwe and is in the final year of an M.Tech. (Hons) degree in Software Engineering at the Harare Institute of Technology (Zimbabwe). Results-driven and innovative teaching assistant in the Department of Software Engineering, with solid academic and practical expertise in software development, systems design, and emerging technologies. She can be contacted at email: stembure@hit.ac.zw.



Wellington S. Manjoro     received the M.Tech. degree in Computer Science and Engineering attained at ITM University Gwalior (India) and also holds a B.Tech. (Hons) degree in Computer Science from the Harare Institute of Technology, Zimbabwe. He has three publications and his areas of interest include intelligent traffic systems, data science, and machine learning among other areas. He is a Microsoft Certified Power BI Data Analyst and holds the Meta Data Analyst Professional Certificate. He can be contacted at email: wmanjoro@hit.ac.zw.