Catching the Spikes: Heteroscedastic Uncertainty Quantification for Enhanced Malaria Prediction

Clark Chen*
New York University
fc2184@nyu.edu

Andrea Qiu*
New York University
Jq2113@nyu.edu

Maggie Qi*
New York University
qq314@nyu.edu

Joyce Zhang*
New York University
kz2085@nyu.edu

Xiang Li†
Peking University
xli2017@nsd.pku.edu.cn

Abstract

Malaria remains a pressing global health challenge and the disease burden is further compounded by fluctuating climatic conditions. To address this problem, researchers explored adapting ConvLSTM and M-LSTM architectures to enhance malaria outbreak forecasting by utilizing key environmental indicators.

However, these models exhibited limitations in accurately capturing sporadic spikes or irregular peaks in malaria outbreak patterns. To conquer this challenge, we utilize maximized log-likelihood of Weibull to develop a new loss function. The Weibull distribution is particularly well-suited for characterizing heavy-tailed, rare events. Our experiment on Pakistan's Malaria outbreak data spanning 2000 to 2017 showed that our new method unlocks the potential of deep learning in public health strategies but also contributes to the advancement of machine learning techniques for handling complex spatiotemporal data with irregular patterns.

1 Introduction

Malaria [Cowman et al., 2016] remains a significant global health challenge, with approximately 249 million cases reported in 2022 by the World Health Organization [WHO, 2023]. The disease's incidence is exacerbated by fluctuating climatic conditions, particularly in equatorial regions. Climate variables substantially influence mosquito breeding cycles and, consequently, the dynamics of malaria transmission, underscoring the importance of sophisticated predictive mechanisms. To address this challenge, researchers have leveraged machine learning to develop robust predictive models [Zinszer et al., 2012, Gaudart et al., 2009]. Specifically, previous work [Nazir et al., 2023] has explored the adaptation of ConvLSTM [Shi et al., 2015] and M-LSTM [Nazir et al., 2023] architectures, renowned for their efficacy in capturing spatial and temporal data dependencies, to enhance malaria outbreak forecasts. These models utilize key environmental indicators such as temperature, rainfall, and Normalized Difference Vegetation Index (NDVI) as predictors, which are crucial for mapping mosquito habitats and potential outbreak zones. Utilizing a comprehensive dataset from Pakistan spanning from 2000 to 2017, encompassing a wide array of climatic and environmental data, these models have been calibrated to predict annual malaria outbreak probabilities with heightened precision.

However, the implementation of prior algorithms revealed their inability to adequately capture sporadic spikes in malaria outbreaks (*Figure* 7). To augment the predictive capability, we employed

---

*New York University

†Peking University

NeurIPS 2024 Workshop on Time Series in the Age of Large Models.

maximum likelihood estimation of the underlying data-generating process, assuming it follows a Weibull distribution. This facilitated the construction of a novel loss function, specifically tailored to enhance model accuracy in predicting outbreak peaks. Our findings not only unlock the potential of deep learning in public health strategies but also contribute to the advancement of machine learning techniques for handling complex spatiotemporal data with irregular patterns.

2 Catching the Spike: Maximum Likelihood of Weibull Distribution

Malaria outbreaks are infrequent occurrences, with the occurrence of such events following a Weibull distribution that captures the rarity of these epidemiological phenomena [Nair et al., 2022], which have wide application in windspeed [Da et al., 2023],dwell times on web pages [Liu et al., 2010] and oil production rate curve of shale oil wells [Lee and Mallick, 2022].. In this study, we introduce a novel approach to predicting disease outbreaks by utilizing the Weibull distribution and incorporating climate data. In contrast to previous work that relies on $\mathcal{L}_1$ loss for model training, our method focuses on learning the entire distribution, by optimizing the scale parameter λ of the Weibull distribution.

In the context of our study, maximum likelihood estimation (MLE) is employed to learn the optimal scale parameter λ of the Weibull distribution, which is used to model the relationship between climate data and disease outbreak rates. By incorporating the Weibull distribution into a neural network architecture and optimizing the log-likelihood function derived from the distribution, we can estimate the value of λ that best fits the observed data, enabling accurate predictions of outbreak rates based on climate conditions.

The likelihood function of a Weibull distribution is given by: $p_k(x|\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$, where $1 > k > 0$ is the shape parameter and $\lambda > 0$ is the scale parameter. In our paper, we assume the outbreak ratio follows a Weibull distribution $p(y|x) = p_k(x|\lambda(x)) = \frac{k}{\lambda(x)}\left(\frac{y}{\lambda(x)}\right)^{k-1} e^{-(y/\lambda(x))^k}$ where $x \in \mathbb{R}^d$ denotes the climate conditions while y is the outbreak rate we observed.

Objective: Maximize the likelihood function, so that the model can learn the optimal scale parameter $\lambda(x)$ of the Weibull distribution. By maximizing the likelihood, we ensure that the learned distribution best explains the observed data, enabling accurate predictions of future outbreaks based on weather conditions. To build the new loss function, we deploy the log-likelihood method. The log-likelihood of Weibull distribution is shown as follows:

$$\log p_k(y|\lambda(x)) = -(y/\lambda(x))^k - k\log\lambda(x) + \underbrace{\log(ky^{k-1})}_{\text{not dependent on thhe prediction } \lambda(x)} \tag{1}$$

We can simplify the full log-likelihood function into following function: $\max_{\lambda(x)} -(y/\lambda(x))^k - k\log\lambda(x) + \underbrace{\log(ky^{k-1})}_{\text{not dependent on parameter } \lambda(x)}$. Since $\log(kx^{k-1})$ is independent term from λ, it does not change as λ changes, thus it can be removed. If we take an inverse of this function and minimize it, it is equivalent to maximize it, which is the goal of the log likelihood. Thus, if our label is y and prediction is $\hat{y}$, we define our loss function as $f(y, \hat{y}) = (y/\hat{y})^k + k\log(\hat{y})$ where y is the label provided in the dataset and $\hat{y}$ is the prediction of our model.

How the Loss Function Learns the Outbreak Rate In this section, we demonstrate that the loss function can learn the outbreak rate y.

Theorem 1. *$f(y, \lambda)$ attains its minimum at $\lambda = y$.*

Proof. The derivative of $f(y, \lambda)$ with respect to λ is: $\frac{\partial f(y,\lambda)}{\partial \lambda} = \frac{\partial (y/\lambda)^k + k\log(\lambda)}{\partial \lambda} = -\frac{ky^k}{\lambda^{k+1}} + \frac{k}{\lambda}$ For λ^* to be a minimum, we must have $\frac{\partial f(y,\lambda^*)}{\partial \lambda^*} = -\frac{ky^k}{\lambda^{*k+1}} + \frac{k}{\lambda^*} = 0$, which implies $\lambda^* = y$. □

Theorem 1 shows that if one minimizes loss the Weibull function, the optimal model he can get will provide exact prediction the label y (outbreak rate in our case). We compare the shape of our loss function with the $\mathcal{L}_1$ loss for different y in *Fig* 1. As illustrated in the figure, Weibull loss can learn the true outbreak rate. Furthermore, Weibull loss will encourage the model to learn a spike and, with larger k values, Weibull loss would converges faster to the true outbreak rate.

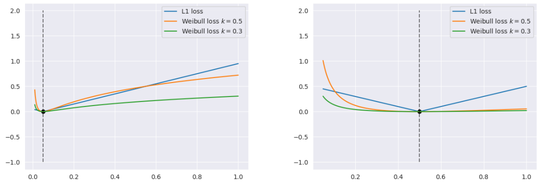

Figure 1: Comparison of Weibull loss and $\mathcal{L}_1$ loss in learning the true outbreak rate (left: 0.05, right: 0.5 (spike)). When the true outbreak rate is not at a spike, the Weibull loss behaves similarly to the $\mathcal{L}_1$ loss. However, when the true outbreak rate is at a spike, the Weibull loss with larger k values converges more quickly to the true rate. This demonstrates its effectiveness in capturing spikes and guiding the model to learn underlying patterns more efficiently than the $\mathcal{L}_1$ loss.

	$\mathcal{L}_1$ **Loss** [Nazir et al., 2023]	**Laplace Loss**	**Weibull Loss** ($k = 0.275$)	**Weibull Loss** ($k = 0.5$)
R^2	0.04	0.09	0.1185	0.113
MAE	0.01	0.0037	0.0027	0.0028

Table 1: Comparasion of $\mathcal{L}_1$ loss, Laplace and Weibull Distribution Testing Loss **ConvLSTM**. The best results are indicated in red and the second best results are indicated in blue.

Rule of the Shape Parameter k A pivotal parameter governing the distribution's behavior is the shape parameter, denoted as k in our paper. Larger values of k correspond to a higher concentration of failures towards the distribution's right tail, effectively skewing the distribution rightward. Consequently, models trained with larger k values exhibit an enhanced propensity for predicting spikes or peaks in the underlying phenomenon. In our study, we trained multiple models employing different k values for the Weibull distribution. Our empirical analysis demonstrated that these models indeed exhibited distinct behaviors in their capacity to predict spikes (*Figure* 6 and 7). Instead of relying on a single optimal k value, which may be challenging to determine, we adopted an ensemble approach. By combining the predictions from these diverse models, our ensemble leveraged the strengths of the individual models while mitigating their respective limitations (*Figure* 2 and 3) .

	$\mathcal{L}_1$ **Loss** [Nazir et al., 2023]	**Weibull Loss** ($k = 0.25$)	**Weibull Loss** ($k = 0.3$)	**Weibull Loss** ($k = 0.35$)	**Weibull Loss** ($k = 0.4$)
R^2	0.117	0.117	0.1669	0.1947	0.286
MAE	0.0029	0.0027	0.0026	0.0027	0.0028

Table 2: Comparison of $\mathcal{L}_1$ loss and Weibull Distribution Testing Loss with different parameters on **M-LSTM**. The best results are indicated in red and the second best results are indicated in blue.

3 Empirical Results

In this section, we implement the Weibull loss with ConvLSTM and M-LSTM and show that our new training algorithm enables the deep learning models to capture the spike structures in the Malaria outbreak data. Our data is derived from the [Nazir et al., 2023], which is based on Pakistan. The data is gathered at distinct locations with specific longitude and latitude coordinate degrees, over each time interval spanning from 2000 to 2017, encompassing 43 distinct features including average NDVI, temperature, and rainfall data for each month. These climate data are collected from the Advancing Research on Nutrition and Agriculture (AReNA) project. Extracting from two distinct satellites, DMSP OLS, and VIIRS, we acquired the average intensity of night-time light data. Also, the dataset has the malaria outbreak rate at each particular spatial point for each year, collected from the Demographic Health Survey(DHS). For the training the dataset was split into two segments: 80% was allocated for training and validation, while the remaining 20% formed the test set. All the models are trained on the same training set and tested on the same testing set.

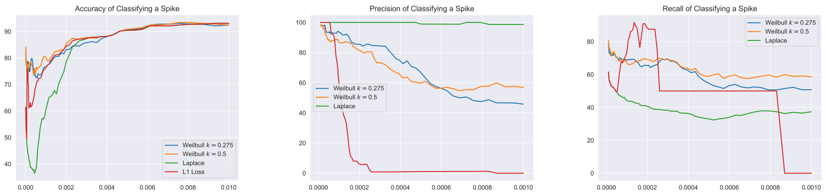

Figure 2: Accuracy, Precision and Recall of generating outbreak warnings using ConvLSTM trained with different loss function. The $x-$axis here is the tolerance set by policy making body to determine if an outbreak has happen. We show that ensembles Weibull model predicts more accurately uniformly with different tolerance.

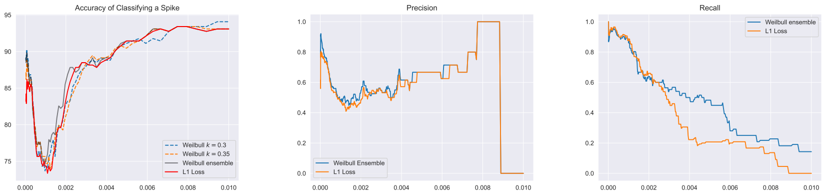

Figure 3: Accuracy, Precision and Recall of generating outbreak warnings using M-LSTM trained with different loss function. The $x-$axis here is the tolerance set by policy making body to determine if an outbreak has happen. We show that ensembles Weibull model predicts more accurately uniformly with different tolerance.

3.1 Evaluation

Table 2 compares the performance of the modified ConvLSTM, which includes $\mathcal{L}_1$ loss, Laplace loss, and Weibull loss with different k values. Generally, the ConvLSTM models trained with the Weibull loss function demonstrate better performance. They show higher R^2 values and lower MAE compared to the Laplace loss and $\mathcal{L}_1$ loss. The best-performing model is the Weibull loss with $k = 0.275$, achieving an R^2 of 0.1185 and an MAE of 0.0027, as indicated in red in Table 1. The second-best model, indicated in blue, is the Weibull loss with $k = 0.5$, which has an R^2 of 0.113 and an MAE of 0.0028.

Figure 2 compares the $\mathcal{L}_1$ loss with the Weibull loss using different k values on the M-LSTM model. The Weibull loss with $k = 0.25$ demonstrates the best performance, with an R^2 of 0.117 and an MAE of 0.0027, as highlighted in red. The second-best model, shown in blue, is the Weibull loss with $k = 0.3$, achieving an R^2 of 0.1669 and an MAE of 0.0026.

In conclusion, the Weibull loss function consistently outperforms the $\mathcal{L}_1$ loss and Laplace loss in both the ConvLSTM and M-LSTM models. The best results are obtained using the Weibull loss with $k = 0.275$ for ConvLSTM and $k = 0.25$ for M-LSTM, demonstrating the effectiveness of the proposed approach in predicting malaria outbreak rates.

	MC-Dropout	**Weibull Loss** ($k = 0.25$)	**Weibull Loss** ($k = 0.4$)
Conv-LSTM	54.2%	79.3%	82.2%
M-LSTM	0.0%	86.5%	92.1%

Table 3: Comparison of coverage using MC-Dropout and Weibull modeling. With large models, MC-Dropout failed to provide meaningful confidence interval while Weibull can still provide reasonable coverage results.

Ability to Classify the Spike and Uncertainty Quantification We further evaluate the ability of our model to classify the spikes. To do this, we recast the evaluation of our model as a classification problem. We consider there is a predefined tolerance tol defined by the government or other authority. If the outbreak rate is larger than the tolerance, then the authority should issue a warning to the public. In this section, we evaluate the accuracy, precision and recall of using different models to generate the warning. As shown in Figure 2 and Figure 3, Weibull loss function can provide uniformly better model with different tolerance compared with the baseline model in [Nair et al., 2022]. In Figure 3, we observe that models trained by the Weibull loss function perform differently depending on the value of k, with certain values of k yielding better performance for high-tolerance scenarios, while other values of k are more well-suited for low-tolerance scenarios. Thus we apply a further ensemble to models trained by the Weibull loss function with different k. The ensemble model generates a warning only when all four models trained by Weibull loss with $k = 0.25, 0.3, 0.35, 0.4$ generate a warning. In Figure 3, we showed that the ensemble model performs better uniformly with all tolerance than the baseline models provided by [Nazir et al., 2023].

Another natural advantage of our framework is enabling uncertainty quantification. For we modeling the outbreak y follows the Weibull distribution $p_k(\cdot|\lambda(x))$, thus we can directly build confidence invertal via computing the quanitle of the Weibull distribution. We plot the 95% confidence interval in Apendix *Figure* 6 and *Figure* 7. We compared the coverage reulsts with standard Monte-Carlo Dropout [Gal and Ghahramani, 2016] used for confidence interval construction for deep learning methods. As shown in *Table* 3, Monte Carlo Dropout is not powerful enough to cover the heavy-tailed rare events while our modeling using Weibul distribution could. Moreover, for larger model such as M-LSTM, Dropout have the ability to overfit the dataset [Liu et al., 2023] which disabled its ability to preform uncertainty quantification.

References

Hoyt Bleakley. Malaria eradication in the americas: A retrospective analysis of childhood exposure. *American Economic Journal. Applied Economics*, 2010.

Cyril Caminade, Sari Kovats, Joacim Rocklov, Adrian M. Tompkins, Andrew P. Morse, Felipe J. Colón-González, Hans Stenlund, Pim Martens, and Simon J. Lloyd. Impact of climate change on global malaria distribution. *Proceedings of the National Academy of Sciences*, 111(9):3286–3291, 2014.

Anthony Costello, Marina Romanello, Stella Hartinger, Georgiana Gordon-Strachan, Saleemul Huq, Peng Gong, Tord Kjellstrom, Paul Ekins, and Hugh Montgomery. Climate change threatens our health and survival within decades. *The Lancet*, 2023.

Alan F Cowman, Julie Healer, Danushka Marapana, and Kevin Marsh. Malaria: biology and disease. *Cell*, 167 (3):610–624, 2016.

Lin Da, Qingshan Yang, Min Liu, Ling Zhao, Teng Wu, and Baolong Chen. Estimation of extreme wind speed based on upcrossing rate of mean wind speeds with weibull distribution. *Journal of Wind Engineering and Industrial Aerodynamics*, 240:105495, 2023.

K. Ebi, A. Capon, P. Berry, C. Broderick, R. Dear, G. Havenith, Y. Honda, R. Kovats, Wei Ma, and Arunima Malik. Hot weather and heat extremes: health risks. *The Lancet*, 398:698–708, 2021.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

Jean Gaudart, Ousmane Touré, Nadine Dessay, A Lassane Dicko, Stéphane Ranque, Loic Forest, Jacques Demongeot, and Ogobara K Doumbo. Modelling malaria incidence with environmental dependency in a locality of sudanese savannah area, mali. *Malaria journal*, 8:1–12, 2009.

Peter W. Gething, David L. Smith, Anand P. Patil, Andrew J. Tatem, Robert W. Snow, and Simon I. Hay. Climate change and the global malaria recession. *Nature*, 465(7296):342–5, 2010.

Se Yoon Lee and Bani K Mallick. Bayesian hierarchical modeling: application towards production results in the eagle ford shale of south texas. *Sankhya B*, 84(1):1–43, 2022.

Chao Liu, Ryen W White, and Susan Dumais. Understanding web browsing behaviors through weibull analysis of dwell time. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 379–386, 2010.

Zhuang Liu, Zhiqiu Xu, Joseph Jin, Zhiqiang Shen, and Trevor Darrell. Dropout reduces underfitting. In *International Conference on Machine Learning*, pages 22233–22248. PMLR, 2023.

Jayakrishnan Nair, Adam Wierman, and Bert Zwart. *The fundamentals of heavy tails: Properties, emergence, and estimation*, volume 53. Cambridge University Press, 2022.

Usman Nazir, Ahzam Ejaz, Muhammad Talha Quddoos, Momin Uppal, and Sara khalid. Towards a spatio-temporal deep learning approach to predict malaria outbreaks using earth observation measurements in south asia. In *NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning*, 2023. URL `https://www.climatechange.ai/papers/neurips2023/66`.

Sachs, Jeffrey, Malaney, and Pia. The economic and social burden of malaria. *Nature*, 2002.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28, 2015.

WHO. World malaria report 2023 — who.int. `https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023`, 2023. [Accessed 09-05-2024].

Kate Zinszer, Aman D Verma, Katia Charland, Timothy F Brewer, John S Brownstein, Zhuoyu Sun, and David L Buckeridge. A scoping review of malaria forecasting: past work and future directions. *BMJ open*, 2(6): e001992, 2012.

Appendix

A Experiment Detail

A.1 Malaria Prediction Data

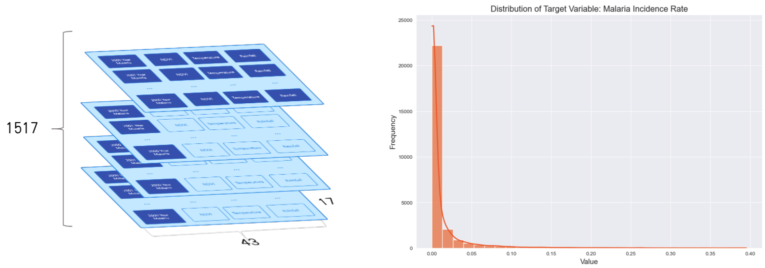

Figure 4: Left Figure: Batch Shape: 43 features (length),17 years (width), 1517 locations (depth). Right Figure: Histogram of original data set. The histogram shows that the distribution of outbreak follows Weibull distribution.

Data Source Our data is derived from the [Nazir et al., 2023], which is based on Pakistan. The data is gathered at distinct locations with specific longitude and latitude coordinate degrees, over each time interval spanning from 2000 to 2017, encompassing 43 distinct features including average NDVI, temperature, and rainfall data for each month. These climate data are collected from the Advancing Research on Nutrition and Agriculture (AReNA) project. Extracting from two distinct satellites, DMSP OLS, and VIIRS, we acquired the average intensity of night-time light data. Also, the dataset has the malaria outbreak rate at each particular spatial point for each year, collected from the Demographic Health Survey(DHS).

Data Cleaning : The central component of our model is the malaria outbreak rate, which serves as the dependent variable, or the targeted y-value. To enhance the efficacy of our modeling process, we normalized this outbreak rate to a range of 0 to 1. We use each 17 years, a particular spatial location as a batch to manage our data.

Training and Testing : In total, we acquired 1517 layers, which were subsequently divided into training and testing sets, with 80% allocated for training and validation tests, and 20% for testing purposes.

A.2 Models

Modified ConvLSTM The ConvLSTM architecture [Shi et al., 2015] typically integrate convolutional structures in both the input-to-state and state-to-state transitions, effectively capturing spatial correlations within the dataset. Specifically, the ConvLSTM employed in the original work utilizes two stacked convolutional layers, each equipped with 3x3 convolutional kernels that preserve spatial dimensions while replacing traditional matrix multiplications find in standard LSTMs. Based on their model, the ConvLSTM operates by processing data from each geographic location in separate batches, allowing for localized feature extraction. After feature extraction through the convolutional layers, the outputs are fed into a standard LSTM layer via a connecting layer. This is followed by a multilayer perceptron that outputs the mean predicted malaria outbreak rates and variance for 2017.

We notice that the original dataset does not align well with their design of ConvLSTM. The specific configurations of their convolutional kernels use the same weights across different features, which do not share an underlying trend. For instance, in the original data, features such as country name and NDVI are sorted as columns without any inherent correlations. This use of a feature-insensitive kernel can not effectively extract meaningful trends between features.

In our project, we make adjustments to our convolutional kernel. We introduce three single layer convolutional neural network equipped with a 1D convolutional kernel (1x12) to extract seasonal patterns of NDVI, temperature, and rainfall from the same year and location separately, and then connected by a MLP. This approach allows us to assign different weights to each feature, enabling more detailed and accurate feature extraction *Fig* 5. These three layers are followed by a multilayer perceptron that connects to the LSTM and outputs the mean predicted malaria outbreak rates and variance for 2017. For Weibull, the output is λ, which can be used to compute mean and confidence interval. Comparison of different loss function using ConvLSTM is presented in Table 2.

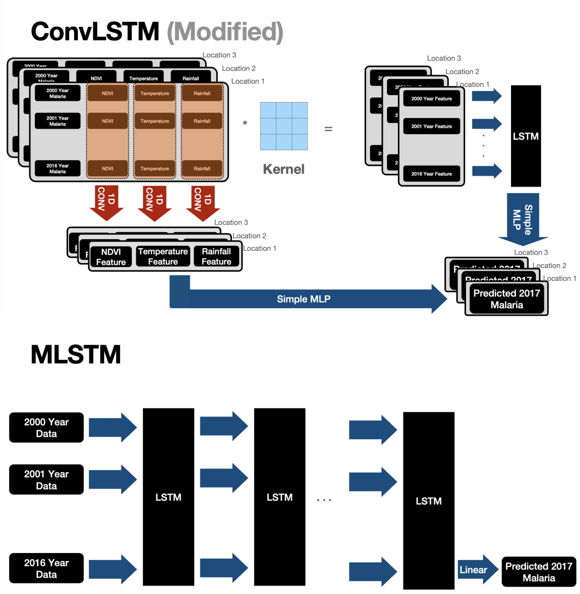

Figure 5: Modified ConvLSTM and M-LSTM Architecture Demonstration

M-LSTM The original work introduces a deep learning methodology termed Multidimensional LSTM (M-LSTM)[Nair et al., 2022], designed to explore the impact of integrating historical environmental and spatiotemporal data in forecasting malaria outbreak rates. At its core, the M-LSTM comprises four LSTM hidden layers, each containing 100 LSTM units designed to capture temporal patterns. These LSTMs are fully connected, culminating in a fully connected linear layer to output the malaria incidence rate for the subsequent year. The model is trained to forecasts the mean malaria outbreak rate and variance of the outbreak for the year of 2017, as illustrated in *Fig* 5. Comparison of different loss function using M-LSTM is presented in Table 2.

B Background on Health Risk of Malaria

Climate change harms health in all populations. Climate change means many infectious diseases expand into previously safe areas and could increase pandemic risk [Ebi et al., 2021]. Many scholars focus on extreme heat events and heat-related health risks [Costello et al., 2023]. This paper focuses on more environmental indicators to capture the not-evenly distributed malaria burden. The global pattern of malarial transmission suggests a disease centred in the tropics, but with a reach into subtropical regions on five continents. Attempts to eliminate or at least suppress the disease have been an important public health story through much of the last century. At malaria's furthest reaches, in temperate zones characterized by strong seasonality and cold winters, these attempts have been successful. Beyond any other factors, this reflects the fact that the base case reproduction rate of malaria is considerably lower in temperate regions than in the tropics, so that moderately intensive efforts at vector control and case management can lead to elimination of the disease. The remarkably high transmission rates in sub-Saharan Africa also reflect the particular capacity of Africa's main vector mosquitoes, the Anopheles gambiae complex of species, with their remarkable tendency towards human biting (anthropophily) [Sachs et al., 2002].

These climatic patterns of course reflect the natural history of the disease. The malaria parasite is transmitted to the female Anopheles mosquito from an infected individual when it takes a blood meal as a prelude to the reproductive process. The parasite must undergo a life-cycle change within the mosquito before it becomes infectious to other individuals in the course of subsequent blood meals. The period required for that life-cycle change increases as the ambient temperature declines, and given the life span of the mosquito, transmission becomes much less likely when the temperature falls below 18°C. Moreover, malaria parasites cease development completely at temperatures below 16°C, and many species of vector mosquitoes suspend biting activity at very low temperatures, further reducing the stability of malaria transmission in temperate regions [Sachs et al., 2002]. Although other climatic features such as rainfall and humidity also affect the stability of transmission, seasonal temperature variation is a predominant factor in explaining the geographical distribution of the disease[Caminade et al., 2014]. Cold winters facilitated effective elimination of malaria infection from much of the temperate zone, leading malariologist Battista Grassi in 1901 to declare malaria a "giant with feet of clay", an obstacle that can readily be eliminated once appropriate interventions become available. With the benefit of hindsight, we now understand that this optimistic statement is at best a statement concerning temperate-zone malaria [Gething et al., 2010]. In tropical regions, exposure to mosquitoes may be perennial and frequently includes several contacts with infected vector mosquitoes each night. Such inoculation rates, combined with the long duration of parasite survival in the host, rapidly saturate local human populations, resulting in universal prevalence and superinfection. This stable pattern of transmission resists amelioration, and vector control efforts that succeed in temperate zones have repeatedly failed to eradicate the parasite from tropical and subtropical regions, although control is possible . The changing global pattern of malaria transmission from 1946 to 1994 illustrates the success of antimalaria efforts in the more temperate regions of the world and the increased concentration of disease burden in the tropics. Today, Africa alone accounts for 90% of malaria mortality [Bleakley, 2010].

Because of the crucial role played by mosquitoes in the transmission cycle, warm and wet climates are more likely to sustain endemic malaria. This pattern implies that malaria outbreaks might result from environmental indicators that are suitable for mosquito survival and reproduction. However, the correlation between environmental indicators and malaria outbreak rates is not linear. This paper captures sporadic spikes or irregular peaks in the malaria outbreak pattern by introducing the Weibull distribution. Predicting these spikes or peaks is of vital importance to slow the spread of the epidemic. This reduces the peak number of people requiring care at one time, so that the health care system does not exceed its capacity.


L1 Loss

Laplace Loss

Weilbull Loss, $k = 0.275$

Weilbull Loss, $k = 0.5$

Figure 6: ConvLSTM with L1, Laplace, Weibull Distribution. The Laplace Model is very noisy. Our Weibull-based Model is able to capture spikes more accurately.

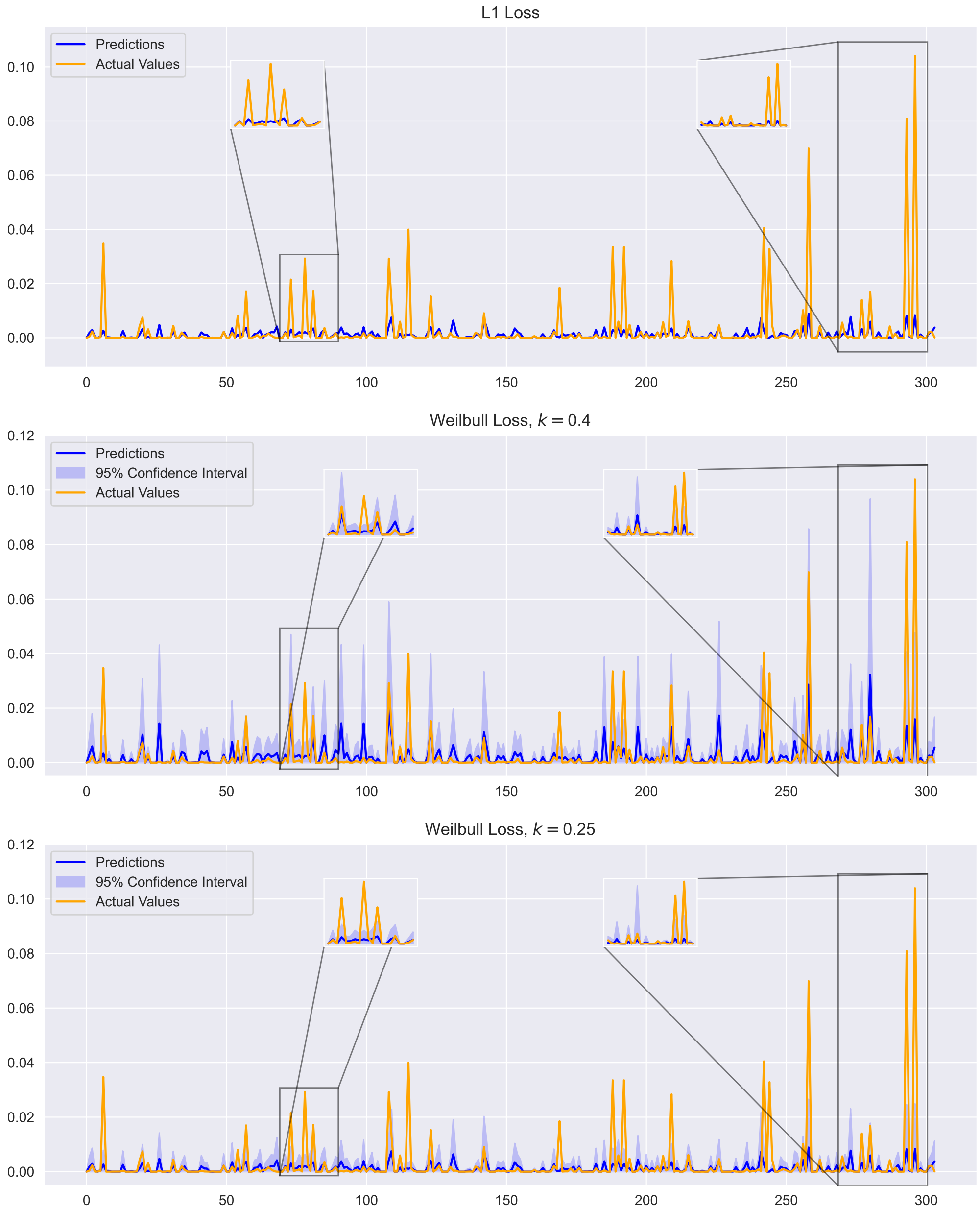

Figure 7: M-LSTM with L1, Laplace, Weibull Distribution. Our Weibull-based Model is able to capture spikes more accurately.

Neurips Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes] ,

   Justification: All the claims in the abstract and introduction is supported by both theorems and experiments.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Yes, we've included limitations including no theoretical coverage guarantee (although showed empirical much higher coverage guarantee).

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: We've provided all the proofs of our theorems.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.

- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We've introduced all the hyperparameter used in our experiment and will open-source our code after our paper is accepted.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We use open source code and the baseline is open-source on github. We'll opensource our code after our paper is accepted.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).

- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We've introduced how our dataset is organized and how we the train/test split.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We've run our paper for five times and report and mean.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We use google colab as our compute resource and all our experiment is easy to reproduce.

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Yes we followed the Neurips Code of Ethics

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We can use our newly trained model to help developing country to predict Malaria and save lives.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not have such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowd-sourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: This paper does not involve crowd sourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.