

Predicting malaria outbreaks using earth observation measurements and spatiotemporal deep learning modelling: a South Asian case study from 2000 to 2017

Usman Nazir, Muhammad Talha Quddoos, Momin Uppal, Sara Khalid



Abstract

Background Malaria remains one the leading communicable causes of death. Approximately half of the world's population is considered at risk of infection, predominantly in African and South Asian countries. Although malaria is preventable, heterogeneity in sociodemographic and environmental risk factors over time and across diverse geographical and climatological regions make outbreak prediction challenging. Data-driven approaches accounting for spatiotemporal variability could offer potential for location-specific early warning tools for malaria.

Methods In this case study, we developed and internally validated a data fusion approach to predict malaria incidence in Pakistan, India, and Bangladesh using geo-referenced environmental factors. For 2000–17, district-level malaria incidence rates for each country were obtained from the US Agency for International Development's Demographic and Health Survey datasets. Environmental factors included average annual temperature, rainfall, and normalised difference vegetation index, obtained from the Advancing Research on Nutrition and Agriculture (known as AReNA) project conducted by the International Food Policy Research Institute in 2020. Data on night-time light intensity was derived from two satellites of the National Oceanic and Atmospheric Administration Defense Meteorological Satellite Program—Operational Linescan System: Nighttime Lights Time Series Version 4, and VIIRS Nighttime Day/Night Band Composites version 1. A multi-dimensional spatiotemporal long short-term memory (M-LSTM) model was developed using data from 2000–16 and internally validated for the year 2017. The M-LSTM model consisted of four hidden layers, each with 100 LSTM units; a fully connected layer was used, followed by linear regression, to predict the malaria incidence rate for 2017 using spatiotemporal partitioning. Model performance was measured using accuracy and root mean squared error. Country-specific models were produced for Bangladesh, India, and Pakistan. Bivariate geospatial heatmaps were produced for a qualitative comparison of univariate environmental factors with malaria rates.

Findings Malaria incidence was predicted with 80·6% accuracy in districts across Pakistan, 76·7% in districts across India, and 99·1% in districts across Bangladesh. The root mean squared error was 7×10^{-4} for Pakistan, $4 \cdot 86 \times 10^{-6}$ for India, and $1 \cdot 32 \times 10^{-5}$ for Bangladesh. Bivariate maps showed an inverse relationship between night-time lights and malaria rates; whereas high malaria rates were found in areas with high temperature, rainfall, and vegetation.

Interpretation Malaria outbreaks could be forecasted using remotely measured environmental factors. Modelling techniques that enable simultaneously forecasting ahead in time as well as across large geographical areas might potentially empower regional decision makers to manage outbreaks early.

Funding NIHR Oxford Biomedical Research Centre Programme and The Higher Education Commission of Pakistan.

Copyright © 2024 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

Contributors

SK, UN, and MU conceived and designed the study. UN and MTQ curated and analysed the data. All authors accessed, interpreted, and verified the data. UN and SK wrote the abstract. All authors revised the abstract. SK is responsible for the overall study. All authors had full access to the data in the study and final responsibility to submit the manuscript for publication.

Declaration of interests

SK is supported by the Innovative Medicines Initiative, Bill & Melinda Gates Foundation, Health Data Research UK, the Alan Turing Institute, British Heart Foundation, and Medical Research Council and Natural Environment Research Council outside of this work. All other authors declare no competing interests.

Published Online

April 15, 2024

Lahore University of Management Sciences, Lahore, Pakistan (U Nazir PhD, M T Quddoos MPhil); Center of Urban Informatics, Technology and Policy, Lahore University of Management Sciences, Lahore, Pakistan (M Uppal PhD); Center for Statistics in Medicine, Botnar Research Center, University of Oxford, Oxford, UK (S Khalid DPhil)

Correspondence to:

Assoc Prof Sara Khalid, Center for Statistics in Medicine, Botnar Research Center, University of Oxford, OX3 7LD, Oxford, UK
sara.khalid@ndorms.ox.ac.uk