Predicting Crimean-Congo Hemorrhagic Fever Outbreaks using Temporal Climate Data
19 July 2022

Presenter: Jonathan Harris
Authors: Jonathan Harris (RPI), Thilanka Munasinghe (RPI), Heidi Tubbs (NASA), Assaf Anyamba (NASA)

Rensselaer Polytechnic Institute
Background

● Crimean-Congo hemorrhagic fever (CCHF)
  ○ Tick-borne infectious disease
  ○ First identified in 1944
  ○ Currently present in Europe, Asia, Africa, and the Mediterranean region [1,2,3,4,5]
  ○ Generally observed between the months of April and September [1]

● Symptoms [6]
  ○ Headache
  ○ Chills
  ○ Nausea
  ○ Vomiting
  ○ High fever
  ○ Abdominal pain
Background

- Transmitted through tick bites or via contact with infected blood or other bodily fluids

- High-risk individuals [7-9]
  - Livestock handlers
  - Slaughterhouse workers
  - Agricultural laborers
  - Family members
  - Healthcare workers
Motivation

● According to NASA,
  ○ Vector-borne diseases (e.g., CCHF) are influenced by trends in climate
  ○ Fatality rate: 3-50%
  ○ Listed as a high-priority disease among international health organizations

● Other works [6,7,10] do not consider the temporal dimension

● If we can better predict CCHF outbreaks using multivariate time-series classification, high-risk countries will have more time to prepare
Related Work

- Other works have developed models to track the spread of tick populations and CCHF using climate predictors
  - Annual environmental sustainability models [10]
    - Analyze effects of climate on tick life cycle and environmental adequacy
  - Boosted regression trees [7]
    - Detection of areas at-risk of CCHF outbreaks
  - Poisson regression [6]
    - Investigate influence of climatic variables on patients

- All of these works identify (in)significant predictive variables and label climate as a key factor
Datasets

● The Program for Monitoring Emerging Diseases (ProMED) Mail Archive (https://promedmail.org/) for CCHF
  ○ Created by Ethan Joseph using a web scraper (https://github.com/sirmammingtonham/vector-borne-disease-analytics)
  ○ Provides alerts of previous CCHF outbreaks around the world

● NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) data repository (https://disc.gsfc.nasa.gov/)
  ○ Provides global precipitation and near surface air temperature data via Giovanni Cloud API [11]
  ○ Specific datasets:
    ■ Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (Half Hourly) [12,13]
    ■ Global Land Data Assimilation System Noah Land Surface Model (3 Hourly) [14,15]
Dataset Construction

- Select a country of interest $C$
  - For training purposes, the ideal country should have numerous reports and various distinct report locations

- Construct dataset $D$ containing the climatic explanatory variables $X_i$ and the binary alert outcome variable $Y$
  - Each row in this dataset is a week in the multivariate time series $X_i$ paired with the alert outcome $Y$ on the day directly after
    - i.e., days 1-7 are paired with the alert outcome of day 8
  - The outcome is 1 if an alert has been reported; 0 otherwise
Country Selection

Reports per Country

- Pakistan
- Russia
- India
- Iran
- Kazakhstan
- Turkey
- Afghanistan
- South Africa
- Uganda
- Serbia
- Namibia
- Iraq
- United Arab Emirates
- Mauritania
- Spain
- Greece
- Oman
- Senegal
- Tajikistan
- Bulgaria
- Republic of Georgia
- Sudan
- Mali
- United Kingdom
- France
- Uzbekistan
- Saudi Arabia
- Canada
- United States
- South Sudan
- Morocco
- Montenegro
- Germany
- DR Congo
- Djibouti
- Denmark
Country Selection

Distinct Locations per Country

- Pakistan
- India
- Russia
- Uganda
- Turkey
- Iraq
- Kazakhstan
- South Africa
- Sudan
- Senegal
- Oman
- Namibia
- Mauritania
- Spain
- Serbia
- Iran
- Afghanistan
- United Arab Emirates
- Mali
- Greece
- Canada
- Uzbekistan
- United States
- United Kingdom
- Tajikistan
- South Sudan
- Saudi Arabia
- Republic of Georgia
- Morocco
- Montenegro
- Germany
- France
- DR Congo
- Djibouti
- Denmark
- Bulgaria
Country Selection

Reports per Distinct Location in Pakistan
Country Selection
Multivariate Time Series Classification

- Model used: MALSTM-FCN [16]
- The dataset $D$ contains under 220 positive examples (3.1%) out of over 7,000 days
- 91.5% classification accuracy

MALSTM-FCN Network Architecture [5]
Future Work

● The incubation period of CCHFV is said to be 14 days so it would be good to experiment with longer time-series sequences for the MALSTM–FCN

● Gather more predictive variables, such as humidity and other climatic factors

● Incorporate another MTSC model, XCM [17], to compare the results

● Test the MALSTM-FCN model with balanced classes
Acknowledgements
Author Profiles

- Jonathan Harris: https://www.linkedin.com/in/jon-harris-7b8296a3/
- Thilanka Munasinghe: https://science.rpi.edu/itws/faculty/thilanka-munasinghe
- Heidi Tubbs: https://science.gsfc.nasa.gov/sed/bio/heidi.c.tubbs
- Assaf Anyamba: https://science.gsfc.nasa.gov/sed/bio/assaf.anyamba
References

References


12. G. J. Huffman, E. F. Stocker, D. T. Bolvin, E. J. Nelkin, and Jackson Tan. 2019. GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06. Greenbelt, MD.


14. H. Beaudoing, M. Rodell, and NASA/GSFC/HSL. 2020. GLDAS Noah Land Surface Model L4 3 hourly 0.25 x 0.25 degree V21. Greenbelt, MD, USA


Thank you!

Questions?