MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

CliZod, compiling the evidence on the climate sensitivity of zoonotic diseases

Emilie Vallee, Masood Sujau, Masako Wada, Teo Susnjak, Barry Borman,
Deborah Read, Hammed Fatoyinbo

Background

- CliZod – Climate and Zoonotic Disease Database aimed at informing climate-sensitive disease modelling
- The project is funded by Wellcome Trust (Digital Technology Development Award in Climate Sensitive Infectious Disease Modelling)
- Consists of a One Health team from Massey University and includes researchers from:
  - Tāwharau Ora - School of Veterinary Science
  - Research Centre for Hauora and Health
  - School of Mathematical and Computational Sciences

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Objective

- Explore how AI methods, especially natural language processing (NLP), can be employed to automate the extraction and compilation of evidence linking climate and disease from scientific literature.
- Seek to deliver a comprehensive insight into the relationship between climate factors and the emergence of zoonotic diseases.
- To serve as a valuable resource for scientists, policy-makers, clinicians, and the global public alike.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Methodology and Approach

- Prioritize diseases and climate parameters

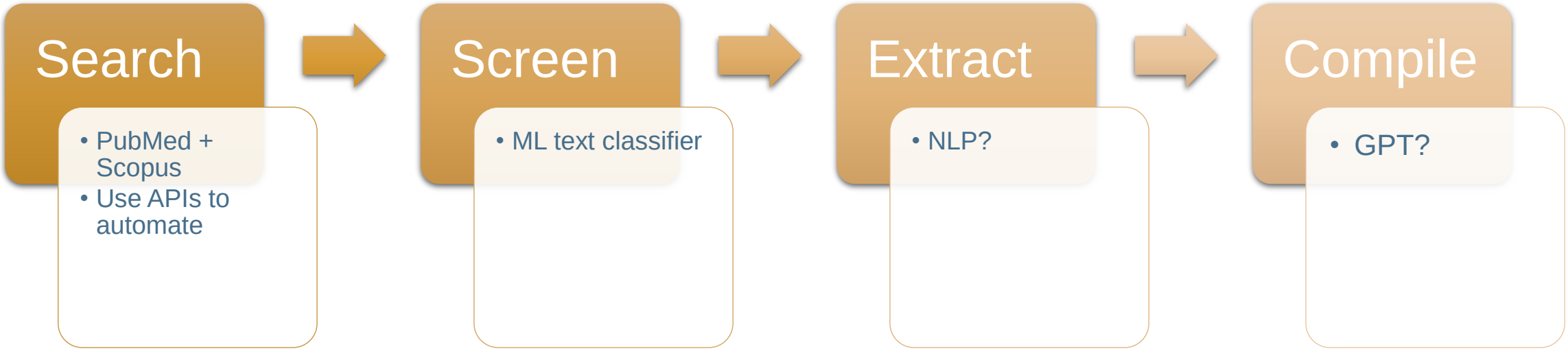
- Searching, screening and scoring abstracts

- Classification

- Parameter extraction

- Synthesis and storage

Methodology and Approach



Prioritize diseases and climate parameters

- Conducted two online questionnaires
  - Identify tools, diseases and variables currently used and considering for future
  - Experience in the field and relevant diseases
  - Demographic information

- Ran workshops for modellers
  - Validate lists of zoonotic diseases and climate parameters
  - Gather modellers' insight data filtering.
  - Identify additional parameters (e.g., species, region) for inclusion.
  - Compile health data sources potential publication insights.

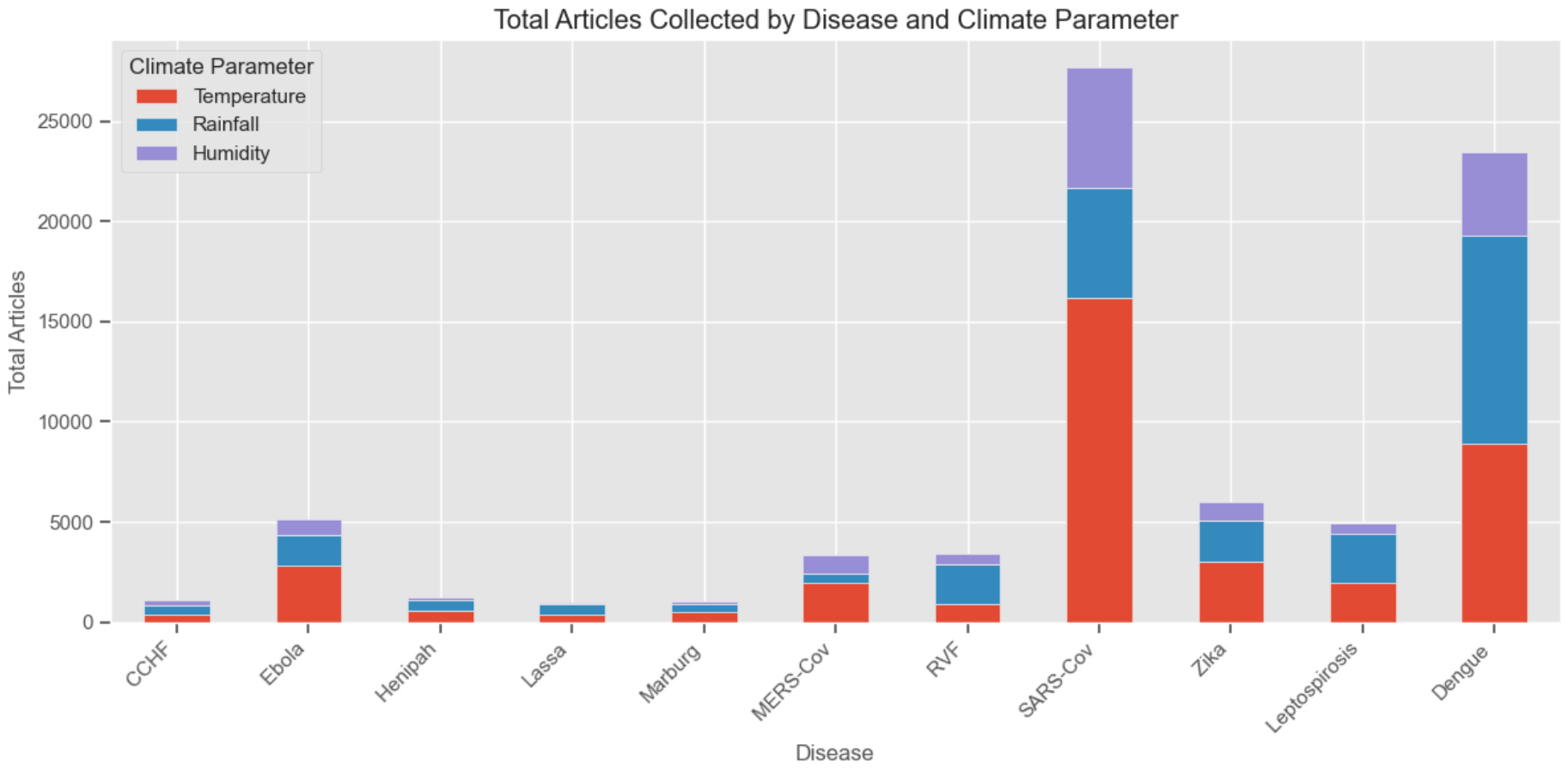
Prioritize diseases and climate parameters

Diseases	Climate Parameters
1. Crimean-Congo hemorrhagic fever	1. Temperature
2. Ebola virus	2. Rainfall
3. Henipah virus	3. Moisture/humidity
4. Lassa mammarenavirus	
5. Marburg virus	
6. Middle East respiratory syndrome coronavirus	
7. Rift Valley fever virus	
8. Severe Acute Respiratory Syndrome coronavirus	
9. Zika virus	
10. Leptospirosis	
11. Dengue	

Searching, screening and classification

- Manual search using PubMed and Scopus
- Searches for 11 diseases and 3 climate variables
- Collecting metadata
- Exploring options for automation
- Established scoring criteria
- Manual scoring to build the training dataset

Training data



MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Classification

- Experimenting with BERT family of models to automate the classification of abstracts based on their relevance to specific diseases and associated climate parameters.
- Using the collected dataset of abstracts along with their manual relevance scores as the training data for the model.
- Aiming to fine-tune the pre-trained BERT model on our dataset, adjusting its parameters to optimize its performance for this task
- Evaluate the performance of the trained model using standard metrics such as accuracy, precision, recall, and F1-score to assess its effectiveness

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Challenges & Questions

- Creating the training dataset
  - ~75,000 abstracts to screen manually
    - How many is enough?
  - ? Articles to read and extract from
- What information do we need to extract? How do we synthesize, given the variety of way to measure and report climate variables and disease occurrence?
  - How do we automate this?
  - Chatbot/GPT to synthesize?
- How to assess the quality of the studies