Imperial College London

Leveraging time series dependencies for clinical management of acute febrile illnesses using machine learning

Damien Ming
d.ming@ic.ac.uk

CAMO-NET meeting
14th December 2023

Imperial College London

Clinical scenarios

- A 59 year old man has been admitted to hospital in the UK with abdominal pain and is awaiting investigations. On day 3 of admission, he develops a fever and becomes less responsive.
- A 24 year old woman develops fever and body aches in Vietnam, and attends the local hospital. Over the next 3 days she develops some gum bleeding but then makes a full recovery.

Imperial College London

Medical observation and patient dynamics

- Change in clinical state over time in the past is important for future prediction
- Methodologically, repeated measurements of different quantities a unique statistical challenge
- Pre-requisites: data of sufficient fidelity, linkage, fusion of data at different frequencies

Imperial College London

Aims of session and discussion:

- Discuss challenges in synthesising time series information for acute care of infections

- Multi-modal incorporation of diverse data (data fusion) and emergent themes

- Implementation of models to the clinical setting

Imperial College London

2 exemplar case studies

Bacterial bloodstream infections (BSI) using blood biomarkers

Dengue severity using pulse waveform and PPG

Imperial College
London

Example 1: Bacterial bloodstream infections (BSI)

- Significant healthcare burden (2.91 million deaths in 2019)
- Positive predictive value of blood cultures ~ 7.5%
- Result turnaround time 24-48 hours
- Priorities in different settings – maximising NPV vs PPV

Imperial College London

Research question

Can the use of longitudinal patient information help predict BSIs and what is the added contribution of such time series data?

Imperial College London

Dataset – iCARE and EHR

- March 2014 – Dec 2021 at Imperial College NHS Trust
- 20,850 patients undergoing blood culture sampling
- Biomarker data up to 14 days prior (where available) extracted

Imperial College London

Microbiological definitions

- BSI state – according to WHO priority pathogens list [1] (18.5%)
- Non-BSI state – no growth at 5 days or isolation only of one of [2] (71.5%)

[1] *Escherichia coli*, *Klebsiella* spp., *Enterococcus* spp., *Pseudomonas* spp., *Proteus* spp., *Serratia* spp., *Citrobacter* spp., *Streptococcus* spp., *Staphylococcus aureus*.

[2] Coagulase negative *Staphylococcus* group, *Micrococcus* spp., *Corynebacterium* spp. excluding *Corynebacterium striatum.*

Imperial College London

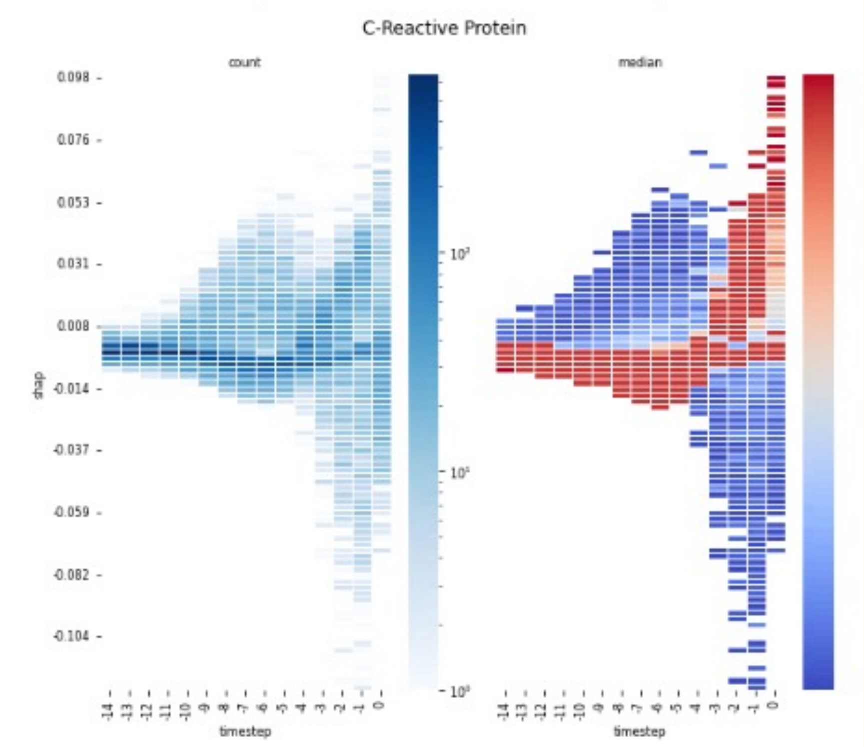
Data methods

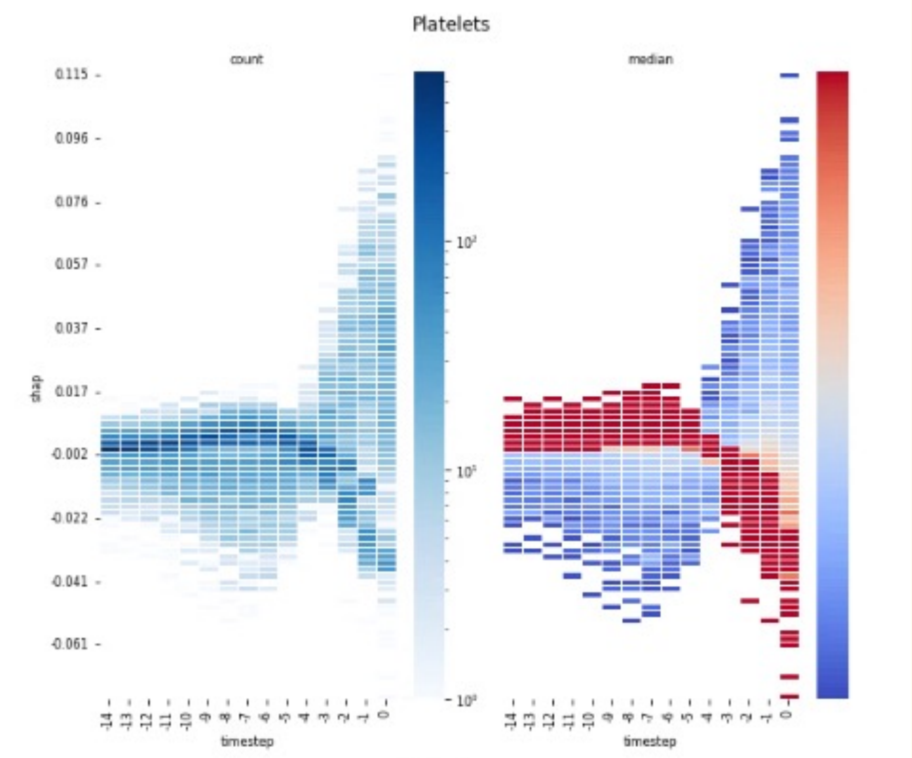
- Initial training and hold-out set temporal split

Training (n=15,212) — March 2014 to Feb 2021

Test (n=5,638) — Feb 2021 to Dec 2021

- 5-fold cross validation
- Missing values forward filled where possible otherwise masked
- A long short-term memory (LSTM) approach for time series data, with logistic regression and feature engineering as baseline
- Tuning using Bayesian optimisation


C-Reactive Protein
count
median
0.098
0.076
0.053
0.031
0.008
-0.014
-0.037
-0.059
-0.082
-0.104
shap
timestep
-14
-13
-12
-11
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0


Platelets
count
median
0.115
0.096
0.076
0.057
0.037
0.017
-0.002
-0.022
-0.041
-0.061
shap
timestep
-14
-13
-12
-11
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0

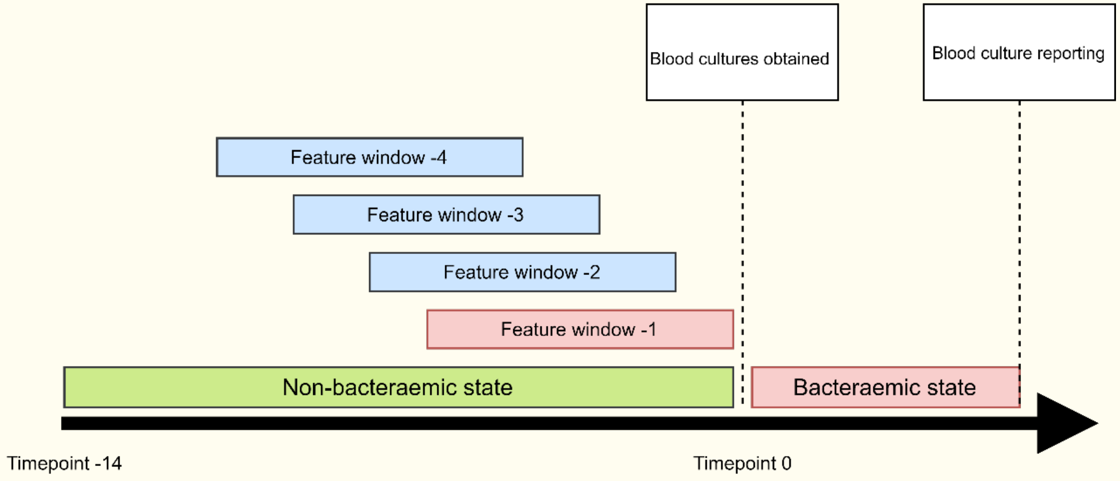
Imperial College London

Handling of missing data and MNAR

- The pattern of biomarker acquisition and measurement is reflective of clinical need
- Forward fill, interpolate, imputation, masking
- Models are built on the premise that local practices remain consistent

Imperial College London

What about a real-time monitoring system?



Imperial College London

Synthesis of metadata to augment prediction

- Patient factors: immunosuppression status, risk factors, clinical assessment derived from natural language processing

- Laboratory metadata: time to positivity vs inoculum volume, information derived from MALID-TOF, use of adjunctive diagnostics e.g. sepsityper

- Fusion of data another challenge…

Imperial College London

Example 2 – dengue risk prediction

- Patients admitted to hospital fulfil dengue with warning signs (and are at increased risk)

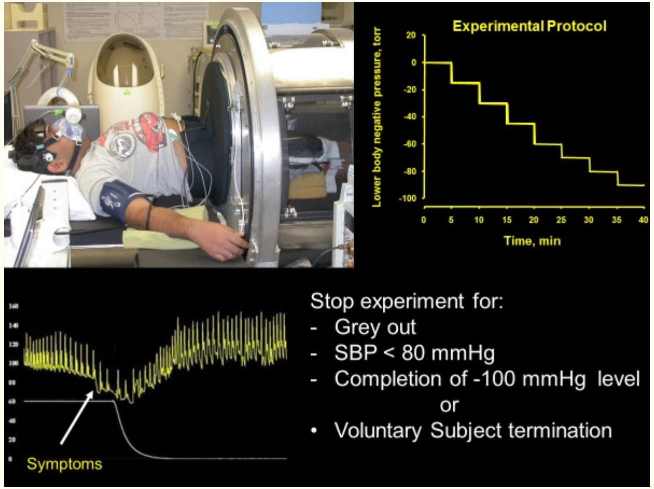
- Tools for close monitoring of patients are lacking

Imperial College London

Wearables using photoplethysmography (PPG)

- Morphology of pulse waveform related to cardiovascular status

(and also derivation of pulse, SpO2, respiratory change)



Imperial College London

Research question

Can pulse waveform alone provide safe patient monitoring in dengue (and prediction of risk in) – and how can its performance be enhanced?

Imperial College London

Methods

- Prospective recruitment of 153 patients admitted to Ho Chi Minh City with acute dengue (2021-2022)

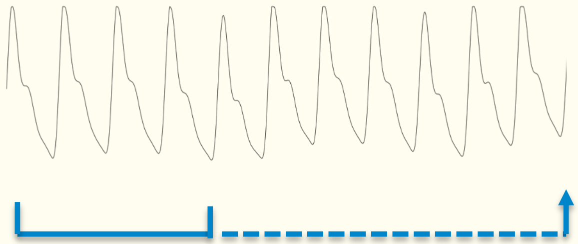
- PPG monitoring for 24+ hours alongside standard vital signs monitoring

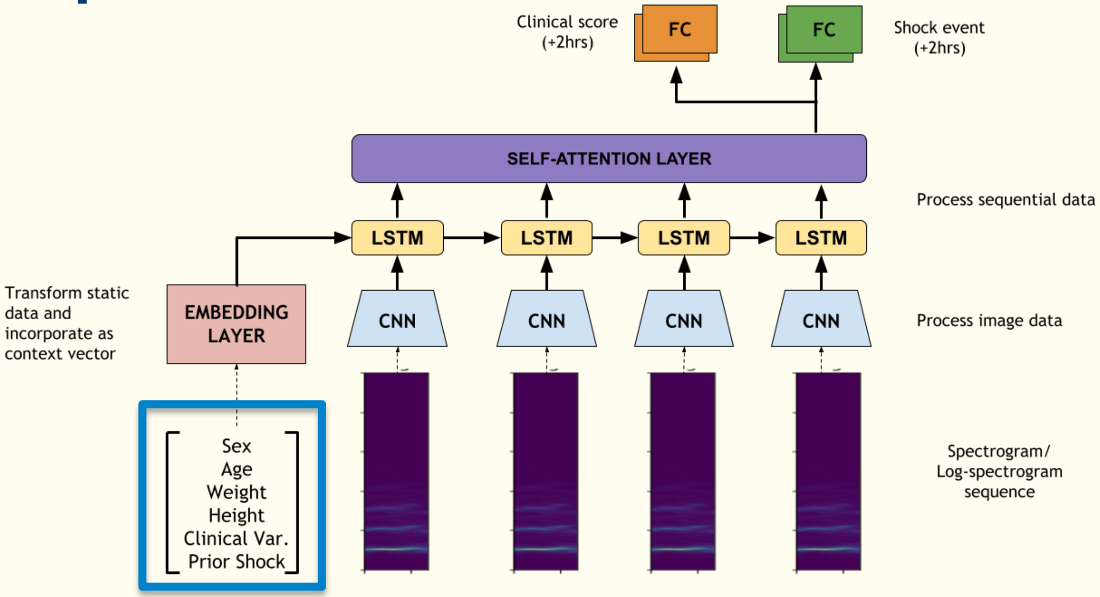
- Approach using PPG signals and prediction of outcomes e.g. NEWS2, dengue shock, other relevant outcomes

- Can we incorporate relevant clinical data to the signal waveform too? Or synthesise with other machine learning models?

Imperial College London

Multi-modal approach to prediction





Dr John Daniels and Professor Pantelis Georgiou
Centre for BioInspired technology, Imperial College London

Imperial College London

Beyond patient related factors

- Synthesis of seasonal, epidemiological, population level data as a proxy of strain virulence etc…

- Dengue prediction and seasonality factors

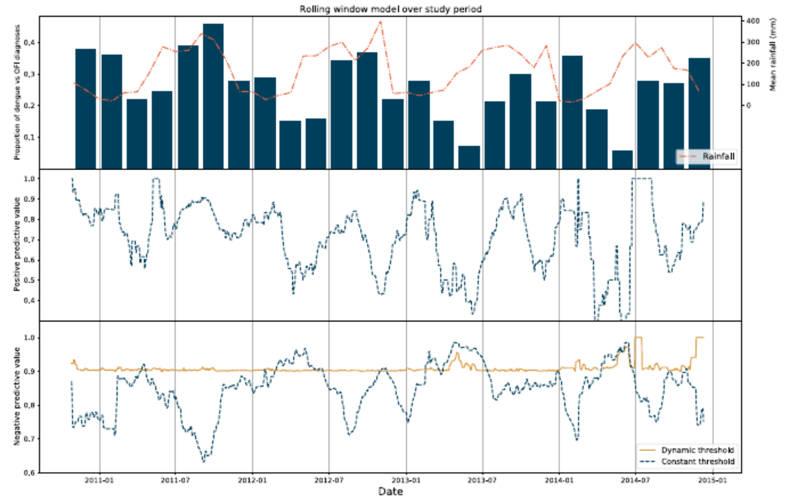


FIGURE 2 | Results of rolling window model. Top: Proportion of dengue diagnoses grouped in 2-month bins plotted against mean rainfall in Vietnam for the entire study period. Middle: Positive predictive value of the rolling window model for the constant threshold model only, with each point on the graph representing the upper range of the 30-day window used as the test set. Bottom: Negative predictive value of the rolling window model for the constant and dynamic model, with the NPV of the latter model fixed to > 90%.

Conclusion

- Time series dependencies are important and a move away from static scores applied at one timepoint could bring benefits to clinical care

- Leveraging such modalities relies on availability of data
  - Connectivity and electronic healthcare records
  - Role of signal acquisition such as through wearable devices

- Complex diverse datasets need synthesis to provide relevant classifications/predictions