

A data-driven approach to predict plasma leakage using explainable machine learning

Ramtin Zargari Marandi^{1*}, Preston Leung¹, Chathurani Sigera², Daniel Dawson Murray¹, Praveen Weeratunga², Deepika Fernando², Chaturaka Rodrigo³, Senaka Rajapakse², Cameron Ross MacPherson¹
 1 - Centre of Excellence for Health, Immunity and Infections (CHIP), Rigshospitalet, Copenhagen University Hospital, Denmark
 2 - Faculty of Medicine, University of Colombo, Sri Lanka
 3 - Viral Immunology Systems Program (VISp), Kirby Institute, UNSW Sydney, Australia

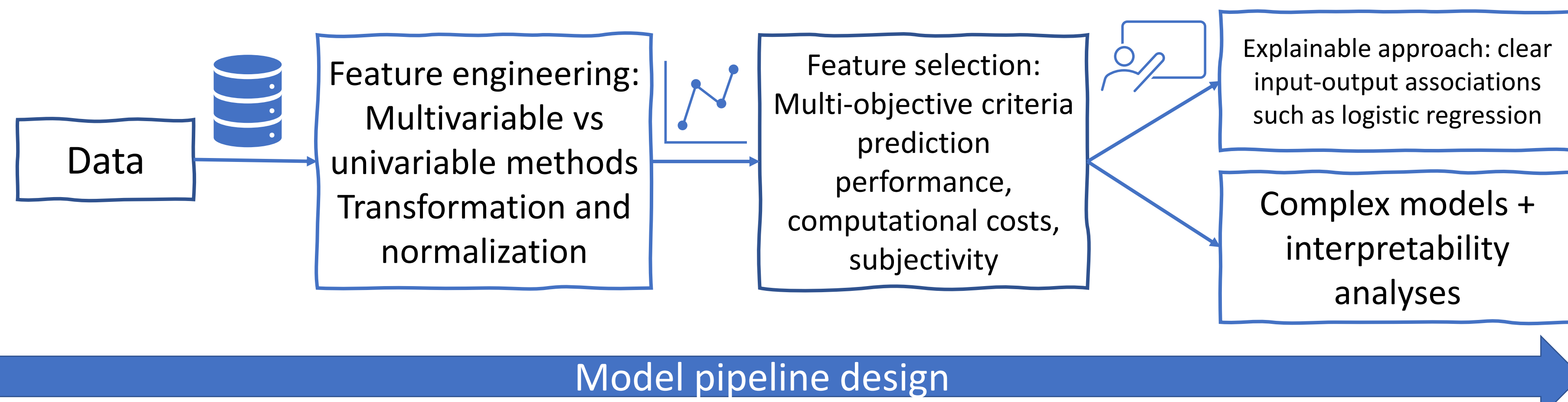


I) Introduction

A decision support system based on a predictive model is vital to save patients in high risk of developing severe complications. This is specially the case in outbreaks where the clinicians need to take a triaging decision on whether a patient should be hospitalized. A certain decision sometimes could not be taken based solely on symptoms and available diagnosis test measurements, but the data collectively could be useful to allow reliable decisions based on predictive models developed by machine learning.

Artificial intelligence has been progressed in different areas specifically in computational platforms allowing researchers with clinical background to implement complex models for disease diagnosis or prognosis based on a wide variety of data. An important challenge is to first formulate the problem based on the data and hypotheses and design machine learning pipeline satisfying multiple important objectives such as interpretability, and cost-efficiency.

In projects with clinical AI as methodology, a primary aim is to develop a model that the clinicians could have clear understanding on how each feature (predictor) contributes to the predictions. This is of high clinical value both on individual level for personalized medicine and on subgroups of patients under the study.



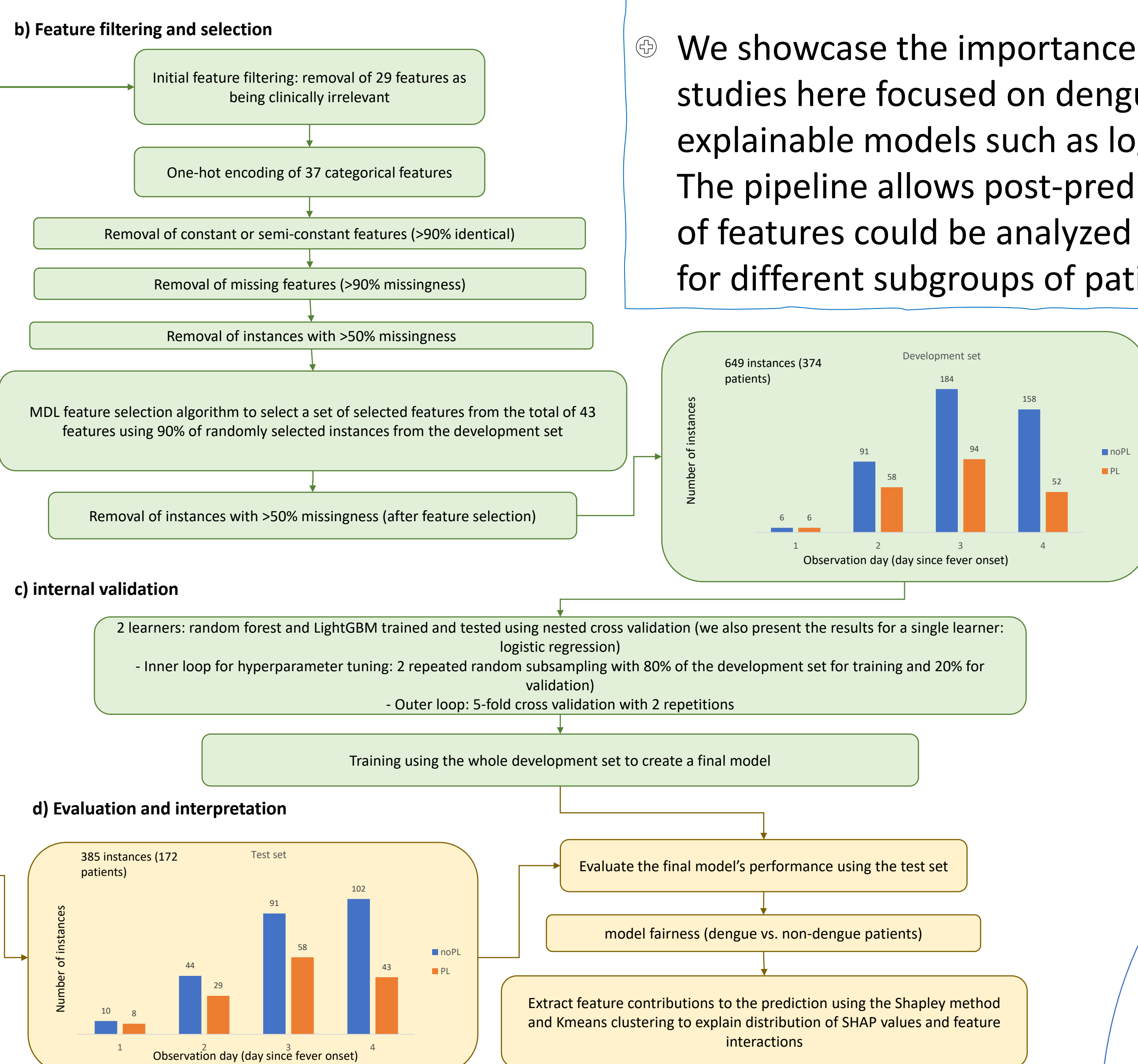
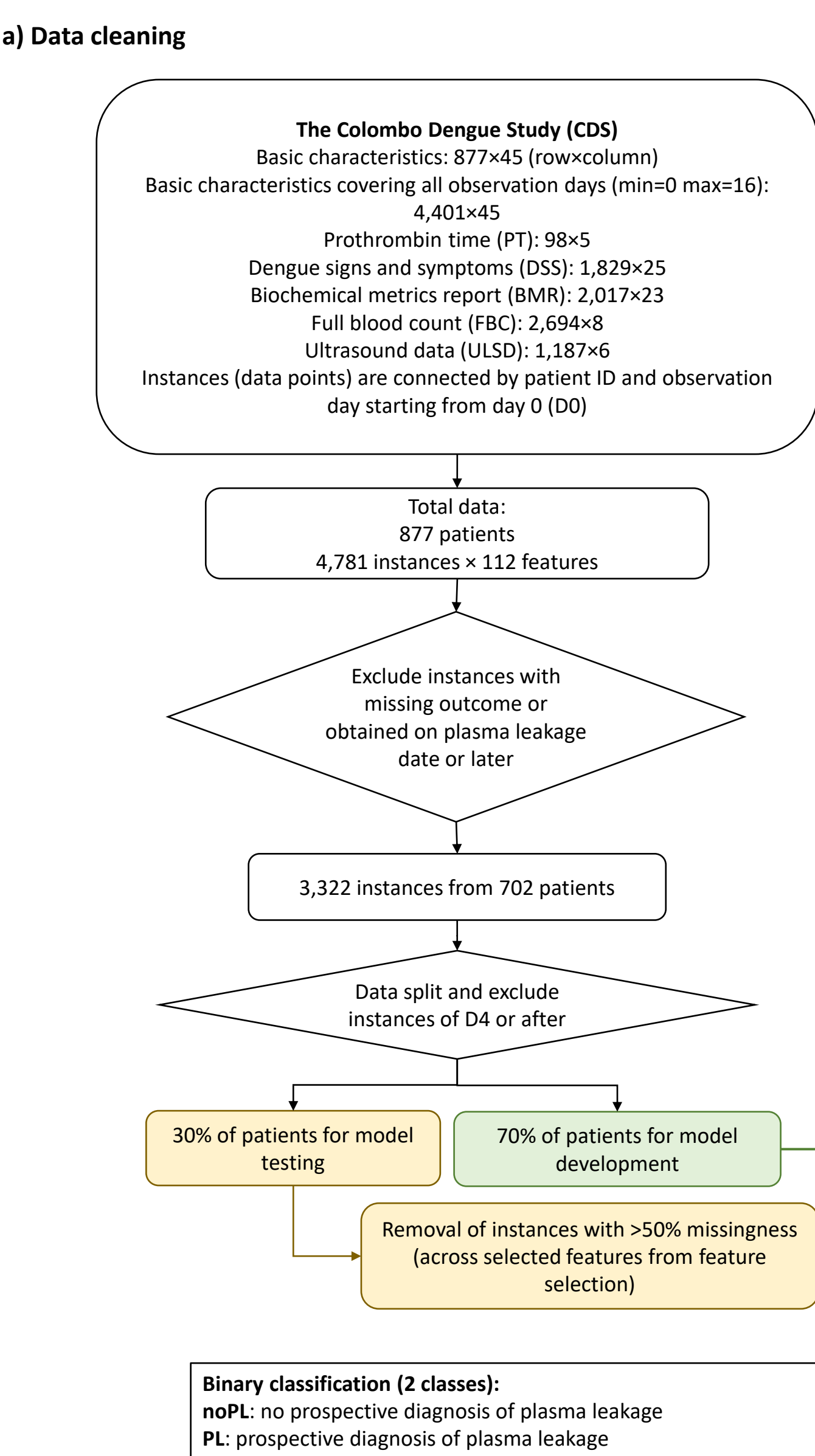
II) Methodology

We used (SHapley Additive exPlanations) SHAP for model interpretability that came from the game theory in which the contribution of each factor (feature or predictor) could be quantified. The method could be computationally expensive in cases where there is a high number of features and samples (instances or patients). Typically, the number of features should be low enough that a human expert could keep track of when interpreting the result of a prediction. We present and discuss quality control steps for data filtering and present useful approach towards more detailed interpretability analyses using SHAP value clustering and interpretability based on correct predictions.

Here we present a real-world case from Sri Lanka where seasonal outbreaks occur due to Dengue infections (a viral infection caused by stings from dengue mosquito). The aim is to identify the patients developing plasma leakage based on routine clinical meta data and blood test measurements in the first four days of patient admission with suspected dengue infection. We present the results from using logistic regression as learner in comparison with more complex models where interpretability is crucial. We also discuss the merits and bottlenecks of both methods.

III) Results

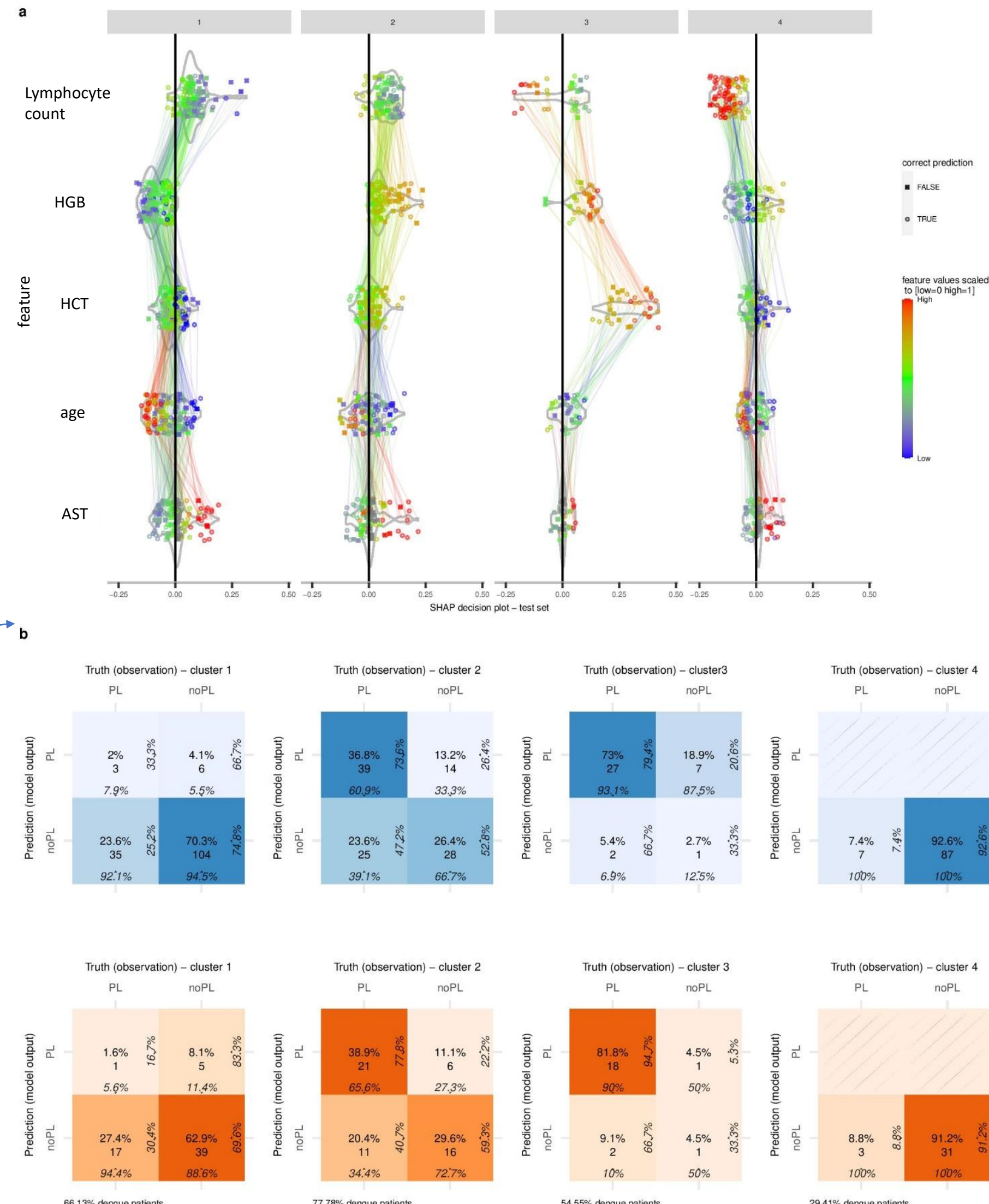
Machine learning pipeline



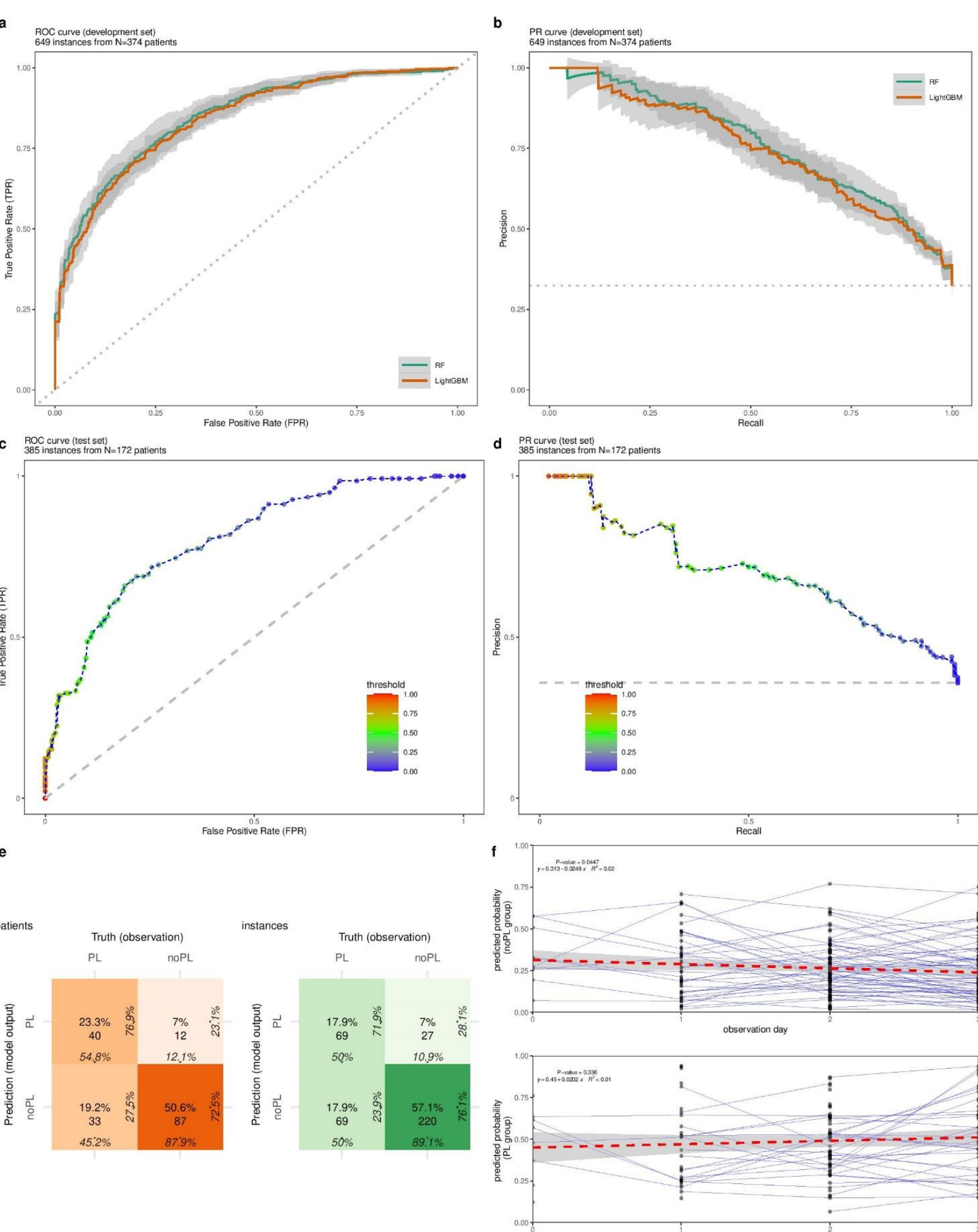
IV) Conclusion

We showcase the importance and challenges of interpretability analyses in clinical AI studies here focused on dengue infection. The presented ML pipeline to utilize explainable models such as logistic regression or more sophisticated ensemble models. The pipeline allows post-prediction analyses for interpretability where the contribution of features could be analyzed on individual level and model usability could be assessed for different subgroups of patients.

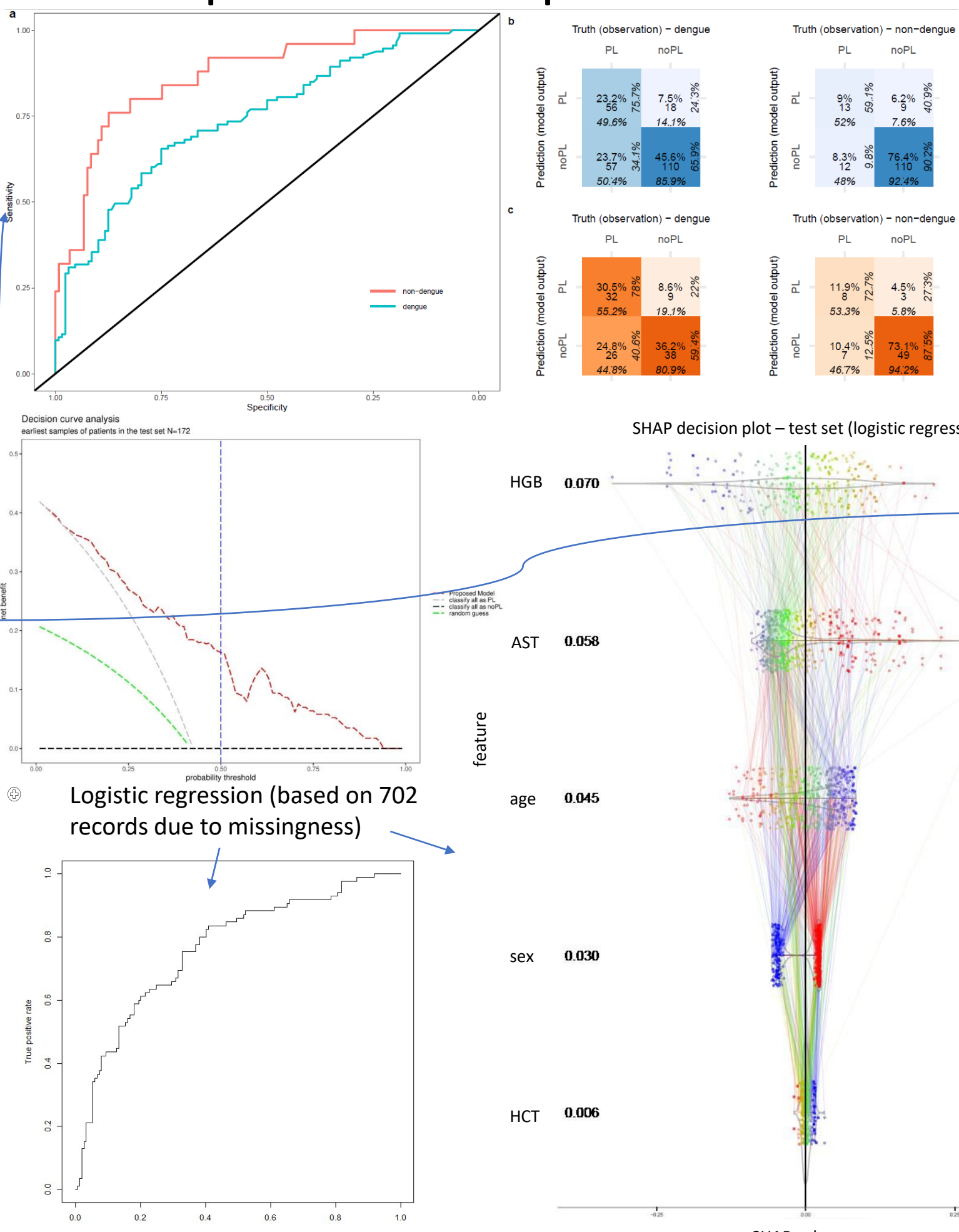
4 SHAP-based clusters describing subgroups of patients with different feature contributions



Model evaluations



Post-prediction steps



V) Acknowledgements

The work was supported by the University of Colombo, Sri Lanka (grant no. AP/3/2/2017/CG/25), the National Health and Medical Research Council, Australia (Investigator grant no. 1173666), and the Danish National Research Foundation (DNRF126).

Find me here:
 R. Z. Marandi, PhD
 @ramtinzm
 ramtin.zargari.marandi@regionh.dk



Let's use low ink cartridges for printing