32nd ECCMID
The congress of ESCMID
Lisbon, Portugal
23–26 April 2022

02005

Learning a meaningful latent space representation for patient risk stratification: model development and validation for dengue

11. Other

#4

B. Hernandez [1, 2], O. Stiff [1], D.K. Ming [2], N.L. Vuong [3, 4], N.M. Tuan [5], N.V.V. Chau [3, 6], H.Q. Chanh [3, 6], N.M. Nguyet [3], N.Q. Huy [3], P.K. Lam [3, 4], D.T.H. Tam [6], D.T. Trung [3], H.T. Trieu [3, 6], C.P. Simmons [7], B. Wills [3], A.H. Holmes [2], S. Yacoub [3], P. Georgiou [1]

[1]Centre for Bio-Inspired Technology, Imperial College London - London (United Kingdom), [2]Centre for Antimicrobial Optimisation, Imperial College London - London (United Kingdom), [3]Oxford University Clinical Research Unit - Ho Chi Minh City (Vietnam), [4]University of Medicine and Pharmacy - Ho Chi Minh City (Vietnam), [5]Children's Hospital 1 - Ho Chi Minh City (Vietnam), [6]Hospital for Tropical Diseases - Ho Chi Minh City (Vietnam), [7]Institute of Vector Borne Disease, Monash University - Clayton (Australia)

Please add any third party affiliations (research group, study group etc.) for your abstract

on behalf of the Vietnam ICU Translational Applications Laboratory (VITAL) investigators

Background

Dengue is a systemic viral disease which exerts a significant health and economic burden worldwide, where up to 5% of hospitalised patients develop life-threatening complications. We evaluated the ability of autoencoders to learn a 2D latent space representation for patient risk stratification which also allows to identify the most likely adverse events.

Methods

Data used consists of an aggregation of prospective clinical studies with children (<18 years old) attending healthcare facilities in Ho Chi Minh City, Vietnam. Autoencoders were used to reduce the dimensionality of the dataset to generate a representative 2D embedding whereby regions are associated with specific clinical phenotypes. Grid Search was used for hyperparameter optimisation using distance metrics to evaluate distance preservation (Pearson and Sheppard diagrams), rank preservation (Spearman) and information loss (Procrusters) and density metrics to evaluate its visualisation (Gaussian Mixture Models ratio).

Results

We included over 9,000 patients recruited between 2000-2021. The latent space produced by the selected autoencoder aligns with established characteristics of the dengue disease progression, such as increase in haematocrit levels, decrease in platelet levels and a

decrease in body temperature (from febrile to critical phase). In addition, similar phenotypes were represented close to each other in the latent space. Balancing distance and density metrics produced results covering most of the latent space, improving visualisation whilst preserving utility for similarity-based retrieval applications, with similar patients grouped closer together. Balance was achieved by using the sigmoid activation function and one hidden layer with three neurons, in addition to the latent dimension layer (Pearson, 0.840; Spearman, 0.830; Procrustes, 0.301; GMM 0.321). Autoencoders with higher number of layers tended to provide good density metric results at the expense of distance metrics and therefore are suitable for other purposes.

Conclusions

This study demonstrates that when adequately configured, autoencoders can produce two-dimensional representations of a complex dataset that conserve the distance relationship between points and therefore group similar patients close together. The visualisation facilitates data interpretation and allows the identification of patients at risk of developing complications. Work is underway to incorporate these findings into an electronic clinical decision support system to guide individual patient management.

Graphical abstract

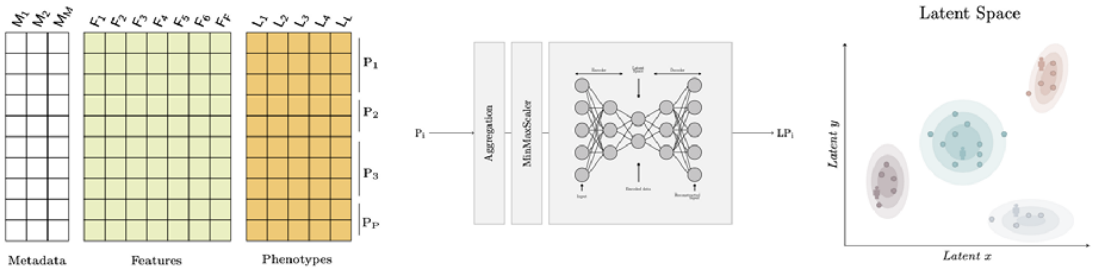

Figure 1: Graphical abstract. On the left, the dataset with metadata, features and phenotypes where each row represents a daily patient profile. In the middle, the model that transforms a patient stay with one or more daily profiles (P_i) into a two dimensional embedding (LP_i) for visualisation. The aggregation step is used to describe the worst patient status using the aggregation functions shown in Table 1. The embeddings are obtained using autoencoders. On the right, the latent space where similar patients are grouped together. Each point represents a patient and the shaded areas represent the density distribution; that is, the concentration of patients for which the phenotype of interest occurs. Note that the latent space can be used to visualise any feature or phenotype of interest.

Latent space analysis

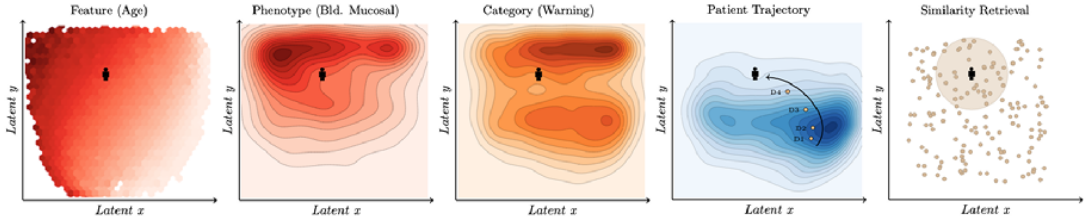

Figure 2: Latent space analysis. From left to right, the latent space produced can be described in terms of features using the average value (e.g. age) and phenotypes (e.g. bleeding mucosal) or categories (e.g. warning signs defined in the WHO 2009 dengue guidelines) using the density distribution. In addition, it is possible to visualise the evolution of the patient over time (patient trajectory) and retrieve previous past similar patients to support decision making (similarity retrieval).

Latent space description: categories

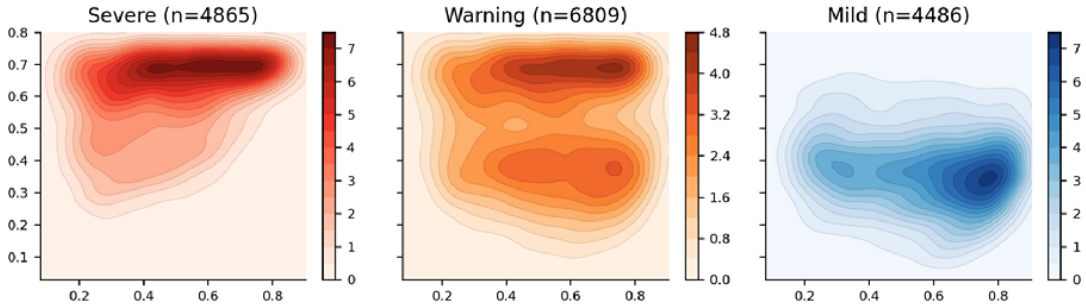

Figure 3: Latent space description: Categories. The graphs represent the density distribution over the latent space for three categories (Severe, Warning and Mild) estimated using a Gaussian kernel.

Conflicts of interest