Imperial College London

Patient risk stratification in dengue with 2D latent space mapping using unsupervised learning

Bernard Hernandez
b.hernandez-perez@imperial.ac.uk
Department of Medicine
Imperial College London
https://bahp.github.io/portfolio-academic/

14th of December 2023

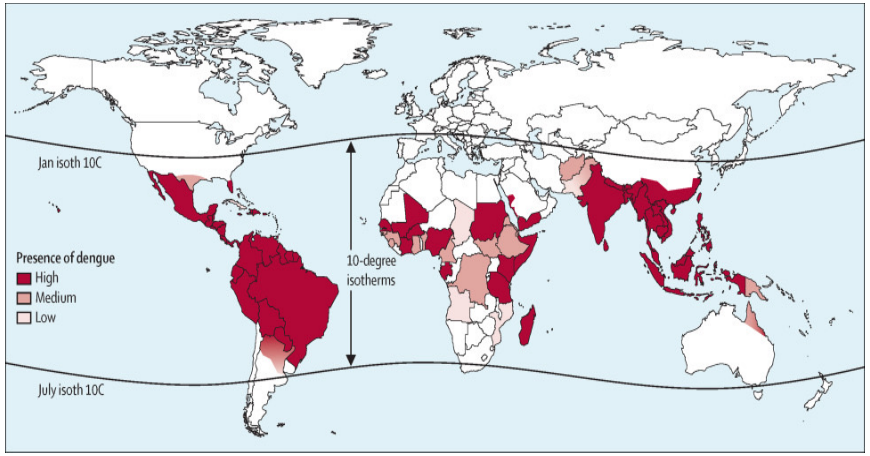
Academic journey

- **Rey Juan Carlos University (URJC), Madrid, Spain**
  - B.Sc. in Telecommunications
  - B.Sc. in Computer Science
- **Royal Institute of Technology (KTH), Stockholm, Sweden**
  - M.Sc. in Machine Learning
- **Imperial College London (ICL), London, United Kingdom**
  - Ph.D. in Computer Science and Healthcare
  - Research Assistant
  - Postdoctoral Research Associate
  - Postdoctoral Research Fellow

A brief overview of Dengue

Introduction

- Dengue fever is a **mosquito-borne** viral infection.
- It is endemic in over 100 countries, with higher prevalence in **tropical** and **subtropical** regions.
- Approximately 51 million symptomatic cases each year, with **seasonal epidemics** and high caseloads imposing a huge strain on local healthcare services.
- Major public health concern with **increasing incidence** (research suggesting climate change influences transmission).

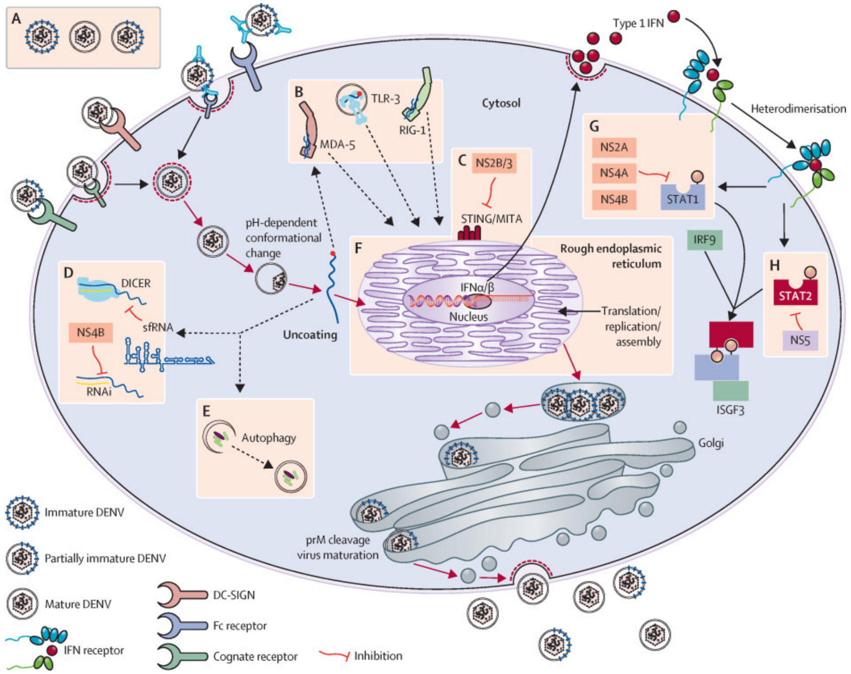


G. Guzman and E. Harris – *Dengue* – The Lancet (2014)

A brief overview of Dengue

Clinical case

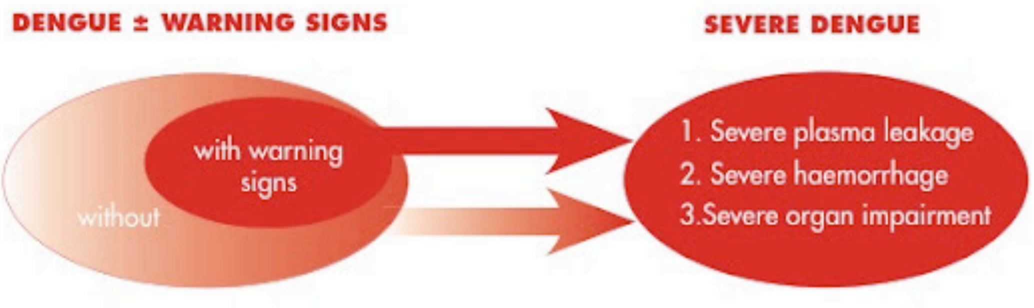
- Wide spectrum and **non-specific clinical symptoms**; potential confusion with other febrile illness.
- Severe cases may develop **bleeding** and **shock** with potentially fatal outcomes.
- Worse outcomes in **children** and young adults.
- Lack of specific antiviral treatment or effective vaccines; management focuses on **supportive care.**



G. Guzman and E. Harris – *Dengue* – The Lancet (2014)

Imperial College London

A brief overview of Dengue



Probable dengue

Live in/travel to endemic area
Fever and 2 of:
- Nausea and/or vomiting
- Rash
- Aches and pains
- Tourniquet test positive
- Leucopenia
- Any warning sign

Warning signs

- Abdominal pain or tenderness
- Persistent vomiting
- Fluid accumulation
- Mucosal bleeding
- Lethargy, restlessness
- Liver enlargement > 2cm
- Haematocrit ↑ & Platelets ↓

Severe dengue

Severe plasma leakage…
- leading to shock (DSS)
- fluid accumulation + distress

Severe bleeding
Severe organ involvement
- Liver: AST or ALT >= 100
- Impaired consciousness (CNS)
- Heart and other organs

World Health Organization – *Dengue: guidelines for diagnosis, treatment, prevention and control.* – *WHO Bulletin* (2009)

Introduction to Case-Based Reasoning

In clinical environments, physician reasoning is based on **knowledge acquired from past** cases personally experienced which is exactly what CBR does! The aim of CBR is to solve new problems based on the solutions of similar past problems in the form of cases

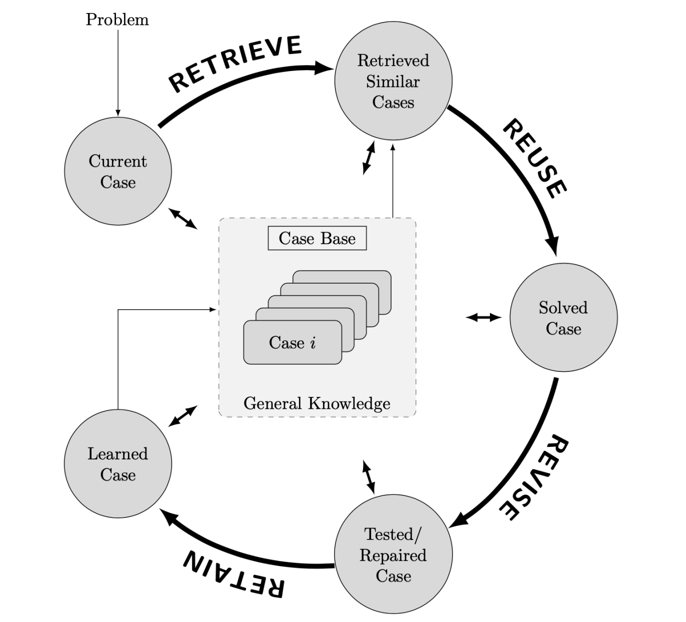
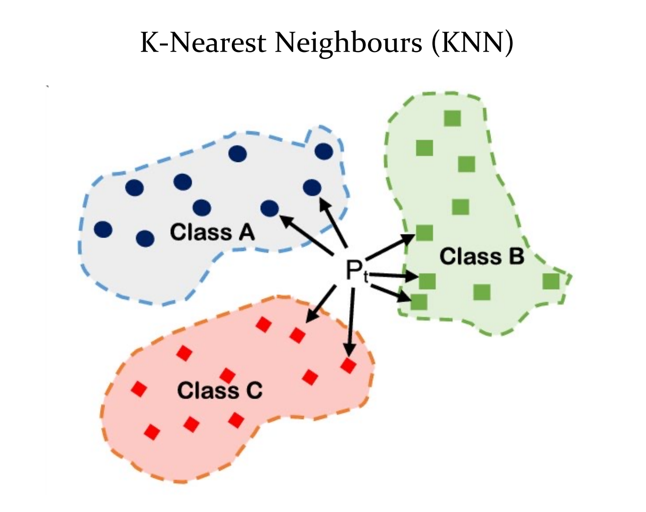
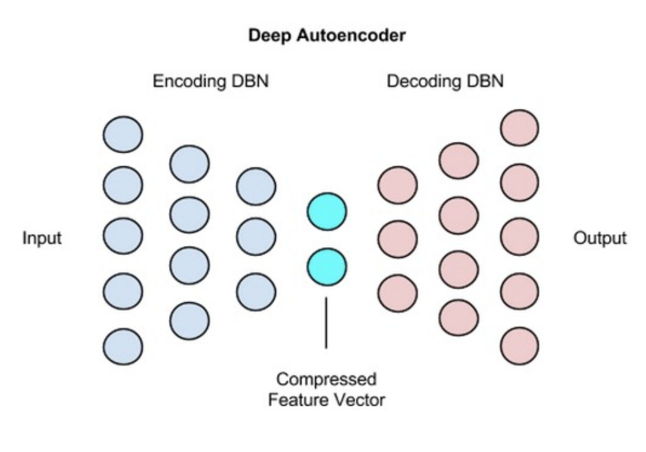
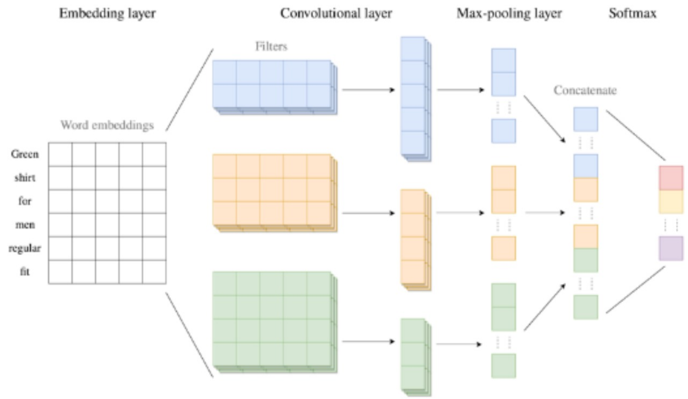


Figure 5.1: The CBR cycle. Diagram showing the different phases for a cycle within the case-based reasoning methodology as outlined by Aamodt and Plaza [7].







1. TM Rawson et al – *A real-world evaluation of a CBR to support antimicrobial prescribing decisions in acute care - CID* (2020)
2. B Hernandez et al – *Data-driven web-based intelligent CDSS: CBR benefits and limitations* – Health informatics (2017)

Problem definition

The dataset used in the study consists of an **aggregation** of prospective clinical data conducted at the Hospital of Tropical Diseases (HTD) and collaborator hospitals in Ho Chi Minh City, **Vietnam** by Oxford University Clinical Research Unit (OUCRU) between **2000 and 2021**.

Code	Year	Population	Type of care	# patients
06DX	2009-2011	A&C	Inpatient	318
13DX	2010-2014	Children	Outpatient	8107
32DX	2013-2016	A&C	Inpatient and ICU	75
42DX	2016-2018	A&C	Inpatient and ICU	664
DF	1999-2009	Children	Inpatient and PICU	1719
DR	2005-2008	Children	Outpatient	1165
FL	2006-2009	A&C	Inpatient	740
MD	2001-2009	Children	Inpatient	3044
01NVA	2020-2021	Children	Inpatient	150*

[1] Only children (under 18 years old) have been considered since they were the most represented in the datasets and there are separate paediatric and adult dengue guidelines.

[2] Dengue diagnosis was defined as one of i) a positive NS1 point of care assay or NS1 ELISA, ii) positive reverse transcriptase-polymerase chain reaction (RT-PCR), iii) positive dengue IgM through acute serology, iv) or seroconversion of paired IgM samples where available.

12,884
patients[1]

19,516
complete daily profiles

4,344 (33.7%)
diagnosed with dengue[2]

9
datasets

20
years

Dataset

Demographics

Gold standard

Aim

1. **Vital-oucru-clinical** documentation (https://bahp.github.io/vital-oucru-clinical/)

Imperial College London

Problem definition

Variable	Value	Overall	A (Severe)	B (Warning)	C (Probable)
n		12884	4866	6809	4486
Age, year		8 [5, 11]	10 [8,13]	8 [5, 11]	6 [3, 9]
Gender	Female	5681 (44.1%)	2180 (44.8%)	3073 (45.1%)	1915 (42.7%)
Weight, Kg		25 [18, 35]	20 [24, 40]	26 [19, 36]	21 [15, 31]
Platelet, k/μL		184 [81, 250]	**68** [36, 143]	**139** [52, 232]	**225** *[177, 277]*
Haematocrit, %		39.8 [36.9, 44]	**45** [41, 49]	**41** [37.7, 46.2]	**37.6** *[35.4, 39.9]*
Body Temperature, °C		37.5 [37.2, 38]	37 [37, 38.4]	37.5 [37.0, 38.0]	37.4 *[37.2, 37.8]*
Mucosal bleeding	True	656 (6.7%)	487 (27.7%)	656 (14%)	-
Vomiting	True	5384 (50.2%)	2499 (90.8%)	5384 (86.9%)	-
Abdominal pain	True	3048 (28.2%)	1686 (60.1%)	3048 (53.3%)	-
Abdominal tenderness	True	1587 (12.6%)	1459 (31.5%)	1442 (22%)	-
Shock	True	1960 (15.2%)	1960 (40.4%)	1700 (25%)	-
Vascular leakage	True	635 (4.9%)	635 (13%)	587 (8.6%)	-
Significant bleeding	True	128 (1%)	128 (2.6%)	118 (1.7%)	-
Organ impairment	True	4709 (36.5%)	4709 (96.8%)	3156 (46.4%)	-
PCR Dengue serotype	DENV-1	1957 (18.1%)	1131 (39.3%)	1305 (22.9%)	373 (8.4%)
	DENV-2	1066 (9.9%)	643 (22.3%)	725 (12.7%)	206 (4.6%)
	DENV-3	321 (3%)	125 (4.3%)	187 (12.7%)	99 (2.2%)
	DENV-4	706 (6.5%)	173 (6%)	381 (6.7%)	297 (6.7%)
	Mixed	18 (0.1%)	18 (0.7%)	12 (0.2%)	-

Dataset

Demographics

Gold standard

Aim

B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Problem definition

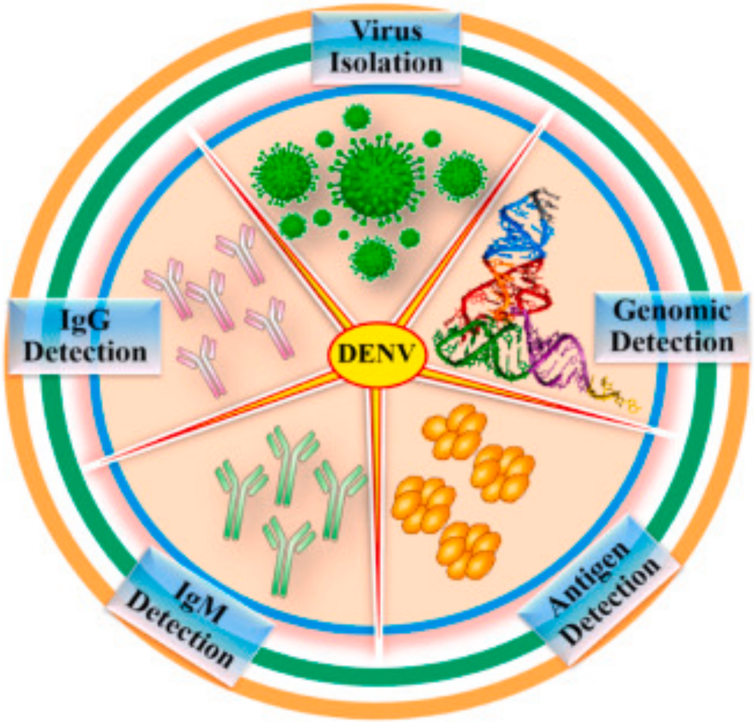
Diagnosis

Polymerase Chain Reaction (PCR)
NS1 Antigen Test
igM positive through acute serology
Seroconversion (ether single or paired igM and igG)

Classification

WHO 2009
- Probable Dengue: Mild
- Warning signs: Medium
- Severe dengue: Severe

WHO 1997
- Dengue Fever (DF): Standard
- Dengue Haemorrhagic Fever (DHF): Medium
- Dengue Shock Syndrome (DSS): Severe



1. S. Subraya et al– *Dengue detection: Advances and challenges in diagnostic technology.* – *Biosensors and Bioelectronics* (2022)
2. World Health Organization – *Dengue: guidelines for diagnosis, treatment, prevention and control.* – *WHO Bulletin* (2009)
3. Repository with vital-oucru-clinical documentation: https://bahp.github.io/vital-oucru-clinical/

Problem definition

AIM: Development of a **clinical decision support** tool to support clinical management of patients with (or under suspicion of) dengue using **unsupervised techniques** to reduce data complexity and facilitate **visualisation**.

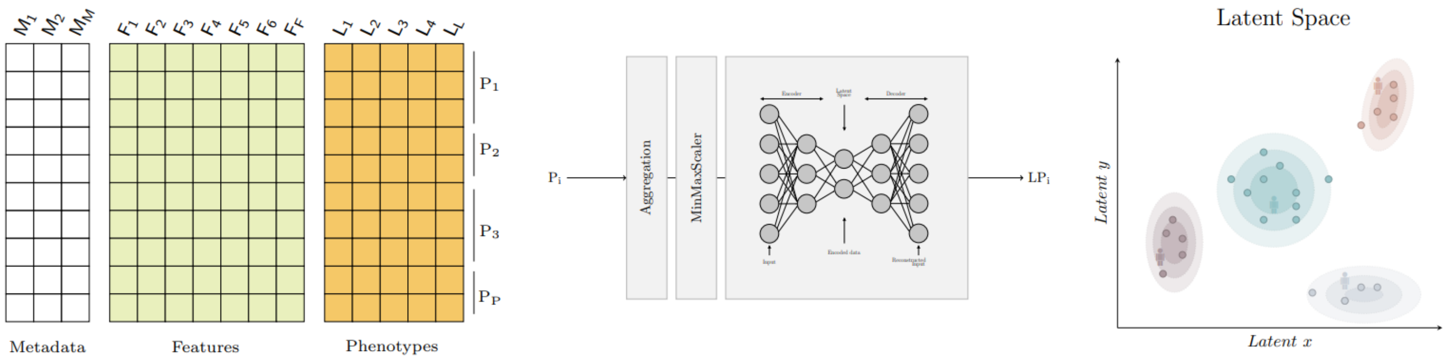


Figure 1: Graphical abstract. On the left, the dataset with metadata, features and phenotypes where each row represents a daily patient profile. In the middle, the model that transforms a patient stay with one or more daily profiles (P_i) into a two dimensional embedding (LP_i) for visualisation. The 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.

Experimental setup

Choosing your **data wisely** is crucial for model **performance** and feasibility of both **implementation** and **adoption**.

Age (year)
Weight (kg)
Body Temperature (°C)
Platelets (k/µL)
Haematocrit (%)

Feature selection

Clinical relevance
Data availability
Quality of data
Frequency of data updates
Cost of data collection
Resource requirements
Turnaround time
Robustness to missing data
Preprocessing requirements
Clinical workflow integration
Interoperability with existing systems
Ethical and Legal considerations
Setting (e.g., LMIC, ICU, …)
Correlation between variables

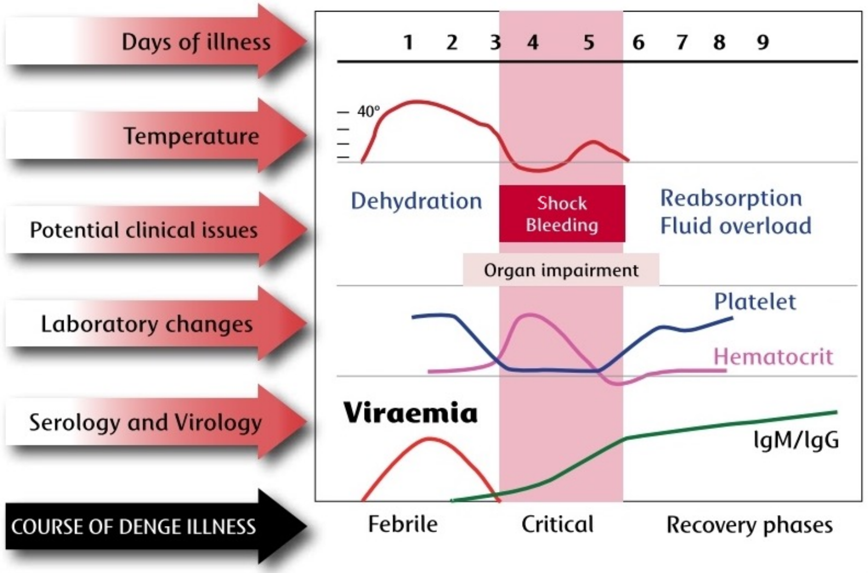


Figure 1. The course of dengue illness diagram. The figure, which has been adapted from WCL Yip, et all 1980 [28], presents phases, lab results, and associated problems.

Experimental setup

Choosing the appropriate **machine learning (ML) approach** according to your data and objectives is crucial for **performance**, **implementation** and **adoption**

- Labels
- Problem type
- Data type
- Complexity
- Interpretability
- Resources
- Scalability
- Adaptation
- Adoption

Algorithm	Type	Continuous space	Unseen data	Comments
PCA	Parametric	✓	✓	The performance is likely to decrease as dimensionality increases due to its linear nature
T-SNE	Non-parametric	✓		Its inability to be applied to unseen data points makes a real-time similarity retrieval system impossible.
UMAP	Non-parametric	✓		Performance is considerably worse than other approaches for dimensionality reduction.
SOM	Non-parametric		✓	The limitations imposed by the discrete space limits the applicability to similarity retrieval.
Autoencoder	Parametric	✓	✓	The ability to encode unseen samples and its support for higher-dimensional data, including time series and images, make it flexible and idea for similarity retrieval.

PCA: Principal Component Analysis; **t-SNE**: t-distributed Stochastic Neighbour Embedding;
UMAP: Uniform Manifold Approximation and Projection; **SOM:** Self-Organising Maps;

Feature selection

Algorithm selection

B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

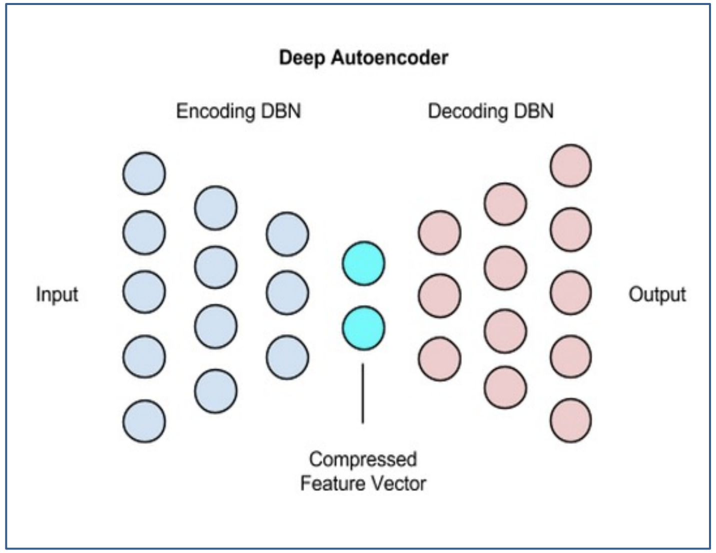
Imperial College London

Experimental setup

Autoencoder: Neural network to encode input data by copying the input to the output. Autoencoders are composed of to elements: **encoder** and **decoder.**

- Minimizes the **difference** between input/output.
- Encourages to learn **meaningful** representations.
- Extracts more important features.
- Works for **different types** of data.

  Such as image, speech, time series, …
- Captures **complex patterns** (non-linear).



Feature selection

Algorithm selection

Autoencoder

B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

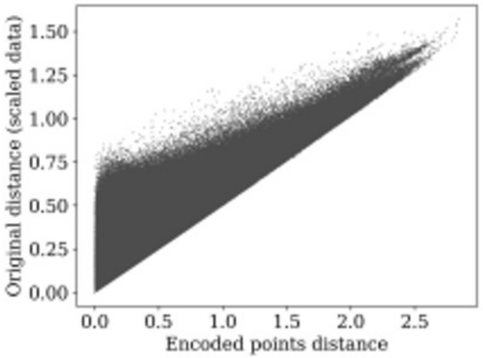
Experimental setup

Table 3
Evaluation metrics.

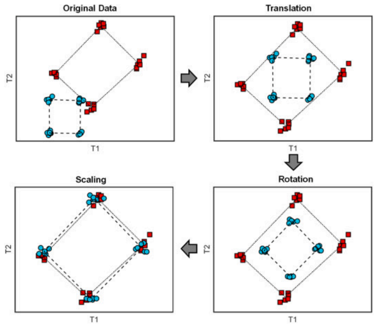
Type	Metric	Aim	
Distance	Sheppard	distance preservation	na
	Pearson	distance preservation	↑
	Spearman	similarity retrieval	↑
	Procrustes	information loss	↓
Density	convex hull ratio	good visualisation	↓
	concave hull ratio	good visualisation	↓
	GMM ratio	good visualisation	↓

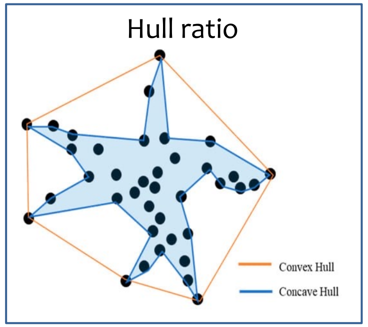
GMM: Gaussian Mixture Model
↑ Higher values are better
↓ Lower values are better

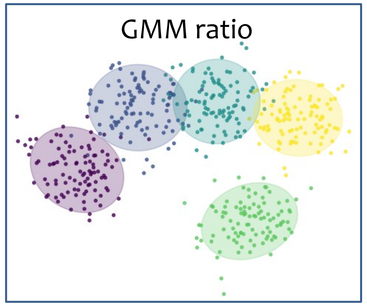
Sheppard



Procrustes







Feature selection

Algorithm selection

Autoencoder

Evaluation

Sheppard: Scatter plot of pairwise distances.
Pearson: Linear correlation to verify if **distances are preserved.**
Spearman: Dependence between rankings, for similarity retrieval.
Procrustes: Compare shape of two objects through superimposition from original to reduced spaces and measures **information loss**.
GMM: Overlapping area within classes. It is more robust to outliers than convex/concave hull ratio.

B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Experimental setup

Cleaning
Unit alignment, spelling and typos, variable names, duplicate data, outlier removal, inconsistent values, format date and time inputs, …

Imbalance
There are **37.7%** of cases diagnose with dengue.
Note we are not doing detection.

Scaling
Used **standard scaling** to help algorithm to converge faster.

Feature engineering
Age – First
Weight – Mean
Body Temperature – Max
Platelets – Min
Hematocrit – Max

Progression of dengue illness follows a **clearly defined** course.
Ability to provide support **from admission**.
Ability to recalculate and **update on every new input**.

12884 patients
19516 daily profiles
Admission on day 2-3…?
Data collection till 7-8…?

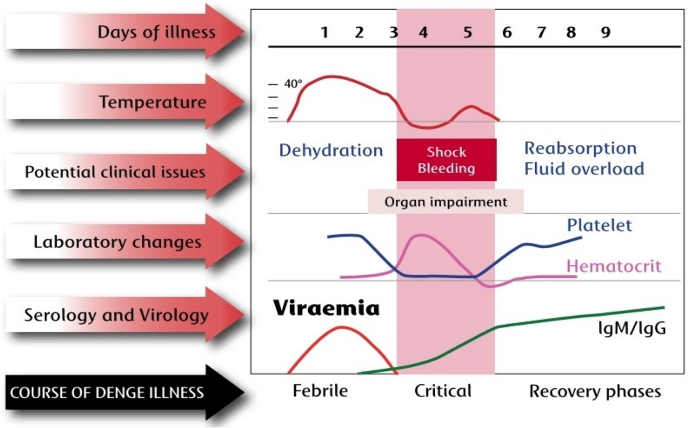


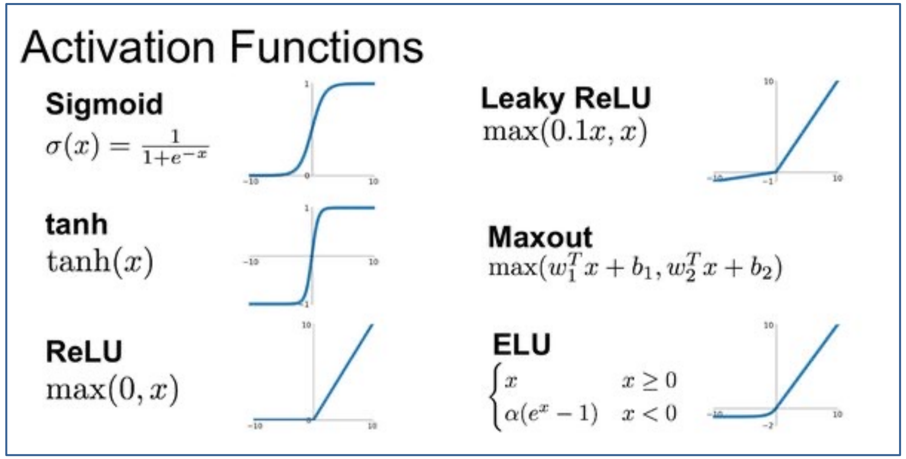
Figure 1. The course of dengue illness diagram. The figure, which has been adapted from WCL Yip, et all 1980 [28], presents phases, lab results, and associated problems.

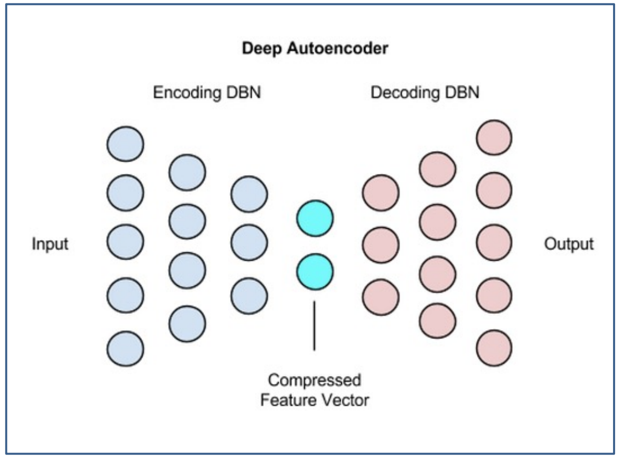
Feature selection
Algorithm selection
Autoencoder
Evaluation
Processing
Cleaning
Imbalance
Imputation
Scaling
Feature Engineering

Experimental setup

Parameter	Comments
Layers[a]	[][b], [5], [4], [3], [5,4], [5,3], [4,3], [5,4,3], [4,4,3,3]
Activation	"ReLU", "Sigmoid"
Learning rate	0.005, 0.001, 5×10^{-4}, 1×10^{-4}, 5×10^{-5}, 1×10^{-5}
Epochs	10, 30, 50, 100, 150, 250, 350, 500
Batch size	16, 32

Table 2. Grid search hyperparameters.
[a] Layers refers to the hidden layers used in the encoder. The input layer and latent layer are not included. The decoder layers are the mirror image of the encoder layers.
[b] No hidden layers other than the latent dimension.



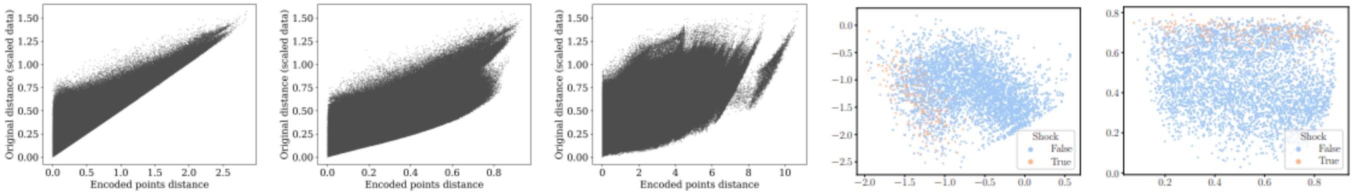


B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Assessment and evaluation

Table 6
Evaluation metrics for various representative hyperparameter configurations

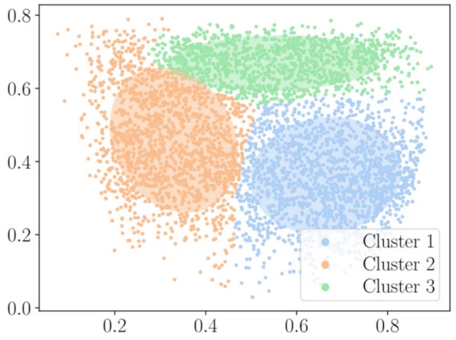
Layers	Activation	Pearson↑	Spearman↑	Procrustes↓	GMM↓	Comments
-	-	0.916	0.896	0.272	0.814	PCA
[ ]	ReLU	0.940	0.920	0.226	0.695	The approximate linearity of the ReLU activation function of this model favours distance preservation.
[ ]	Sigmoid	0.917	0.906	0.240	0.543	The non-linearity of the Sigmoid activation affects distance metrics slightly and improves density metrics.
[3]	Sigmoid	0.840	0.830	0.301	0.321	It balances distance preservation and density metric results.
[5,4,3]	ReLU	0.635	0.622	0.505	0.104	It is a complex model with good density metric results but produces dense points in the latent dimension not apt for visualisation of patient trajectories over time. In addition, distance metric results show that distances are not preserved and therefore it is inadequate for similarity-based retrieval.



(a) () ReLU **(b)** (3) Sigmoid **(c)** (5,4,3) ReLU **(d)** (3) ReLU **(e)** (3) Sigmoid

Figure 3: Sheppard diagrams (left) and shock label projections (right). On the left, Sheppard diagrams obtained for autoencoders with three different configurations. On the right, distribution of patients in the latent space with (orange) and without (blue) shock.

Assessment and evaluation



		Overall	Cluster 1	Cluster 2	Cluster 3
n		14484	5588	5017	3879
abdominal_pain, n (%)	False	9878 (68.2)	4533 (81.1)	3780 (75.3)	1565 (40.3)
	True	4606 (31.8)	1055 (18.9)	1237 (24.7)	2314 (59.7)
ascites, n (%)	False	12153 (83.9)	5147 (92.1)	4001 (79.7)	3005 (77.5)
	True	2331 (16.1)	441 (7.9)	1016 (20.3)	874 (22.5)
bleeding, n (%)	False	10760 (74.3)	5133 (91.9)	3831 (76.4)	1796 (46.3)
	True	3724 (25.7)	455 (8.1)	1186 (23.6)	2083 (53.7)
bleeding_gum, n (%)	False	12895 (89.0)	4844 (86.7)	4372 (87.1)	3679 (94.8)
	True	1589 (11.0)	744 (13.3)	645 (12.9)	200 (5.2)
bleeding_mucosal, n (%)	False	11818 (81.6)	5387 (96.4)	4432 (88.3)	1999 (51.5)
	True	2666 (18.4)	201 (3.6)	585 (11.7)	1880 (48.5)
bleeding_skin, n (%)	False	7864 (54.3)	4820 (86.3)	2490 (49.6)	554 (14.3)
	True	6620 (45.7)	768 (13.7)	2527 (50.4)	3325 (85.7)
gender, n (%)	Female	6327 (43.7)	2563 (45.9)	1922 (38.3)	1842 (47.5)
	Male	8157 (56.3)	3025 (54.1)	3095 (61.7)	2037 (52.5)
shock, n (%)	False	13783 (95.2)	5576 (99.8)	4905 (97.8)	3302 (85.1)
	True	701 (4.8)	12 (0.2)	112 (2.2)	577 (14.9)
age, median [Q1,Q3]*		8.0 [5.0,11.0]	4.0 [3.0,6.0]	11.0 [10.0,13.0]	10.0 [8.0,12.0]
temperature, median [Q1,Q3]*		37.6 [37.2,38.3]	37.4 [37.2,37.8]	37.9 [37.4,38.5]	38.0 [37.0,38.8]
hct, median [Q1,Q3]*		40.3 [37.2,45.0]	37.2 [35.1,39.2]	41.0 [38.6,44.0]	46.4 [43.0,50.0]
plt, median [Q1,Q3]*		169.0 [71.0,243.0]	229.0 [182.0,279.0]	182.0 [109.0,243.0]	46.0 [30.0,69.0]
weight, median [Q1,Q3]*		26.0 [19.0,37.0]	18.0 [14.0,22.0]	38.0 [30.0,46.0]	29.0 [22.0,36.0]

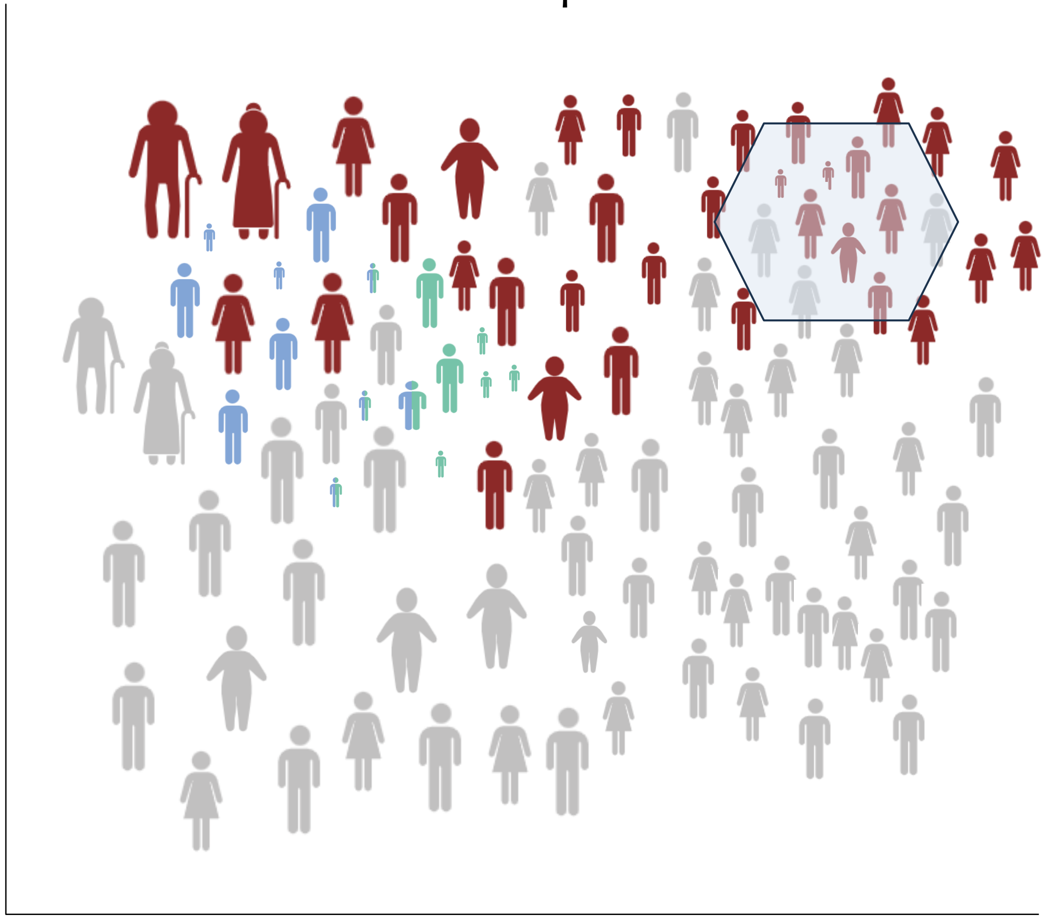
Performance

Clustering

O. Stiff– *Patient similarity retrieval based on laboratory data cluster analysis and visualization* – *M.Sc. thesis* (2022)

Assessment and evaluation

Latent space



- Query patient
- Male / phenotype A
- Female / no phenotype A
- Obesity
- Older patient
- Hexagonal bin
- Phenotype B
- Phenotype C
- Phenotype B & C

Performance

Clustering

Further analysis

Feature (e.g. Age)

= mean (12*3 + 11*2 + 9*4 + 3*1) = 9.7

Phenotype (e.g. Shock) *

≈ 7 / (7 + 3) = 0.7 = 70%

Category (e.g. Female & Phenotype) *

≈ 2 / (2 + 8) = 0.8 = 80%

* It is actually the density distribution using Gaussian Kernel

B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Assessment and evaluation

Each point on the graph represents a complete patient's hospital stay, with those experiencing **shock** displayed in orange. They all seem to be grouped on the **top area**.

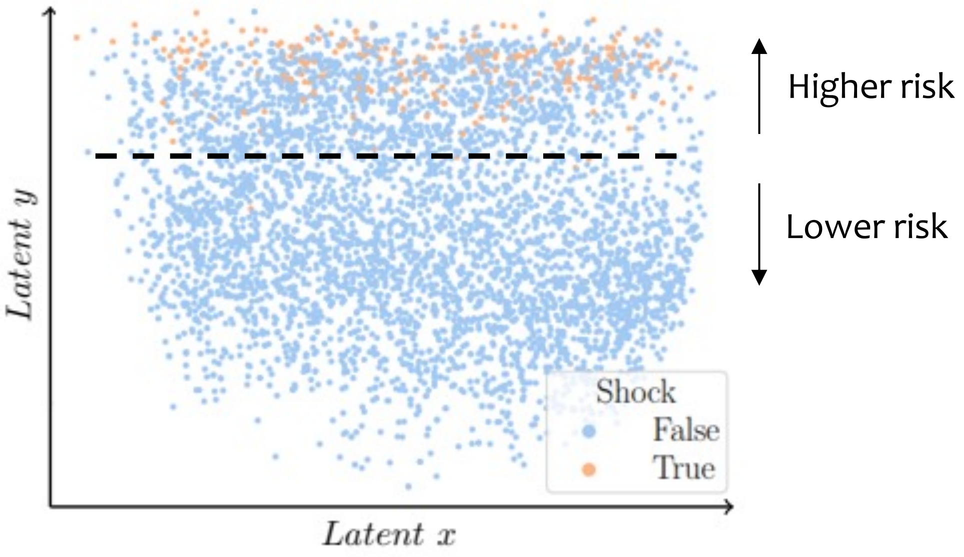


Figure 2. Latent space: embeddings. The worst patient status for all the patients has been projected into the latent space with the shock phenotype.

B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Assessment and evaluation

Aligns with established **characteristics of disease progression** such as increase in haematocrit levels ↑, decrease in platelet levels ↓ and decrease in body temperature from febrile to critical phase.

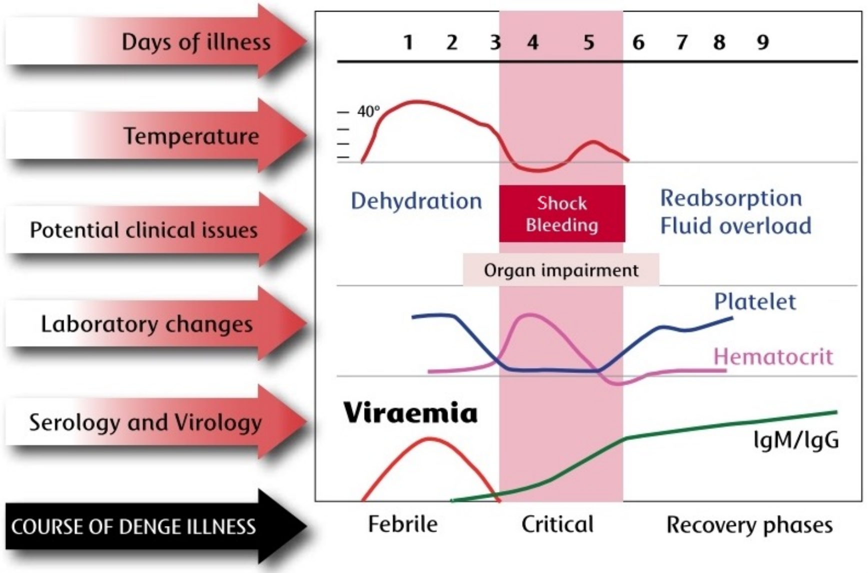


Figure 1. The course of dengue illness diagram. The figure, which has been adapted from WCL Yip, et all 1980 [28], presents phases, lab results, and associated problems.

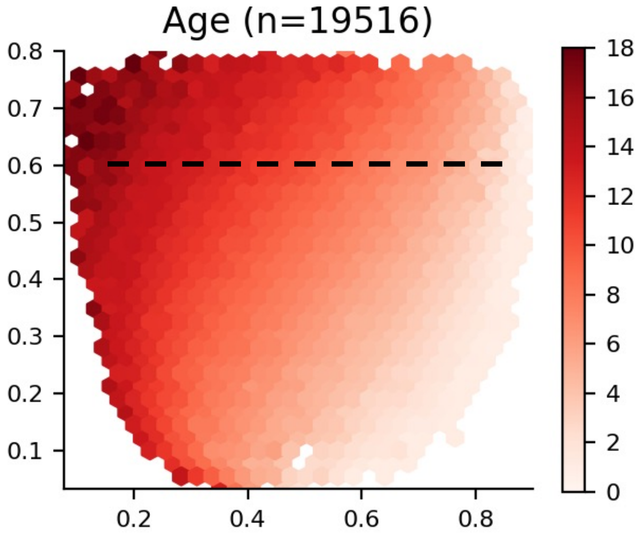


Figure 3. Latent space description: Features. The graphs represent the density distribution using hexagonal binning over the latent space.

Assessment and evaluation

Aligns with established **characteristics of disease progression** such as increase in haematocrit levels ↑, decrease in platelet levels ↓ and decrease in body temperature from febrile to critical phase.

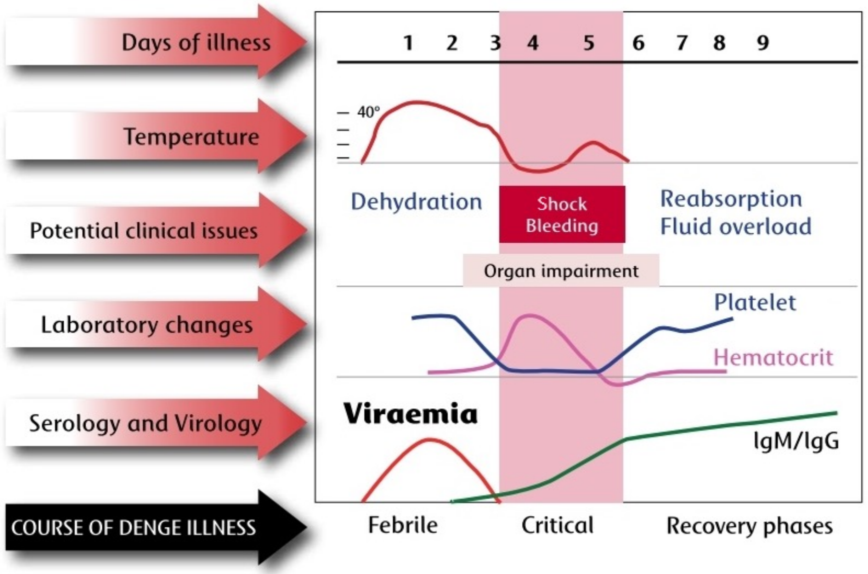


Figure 1. The course of dengue illness diagram. The figure, which has been adapted from WCL Yip, et all 1980 [28], presents phases, lab results, and associated problems.

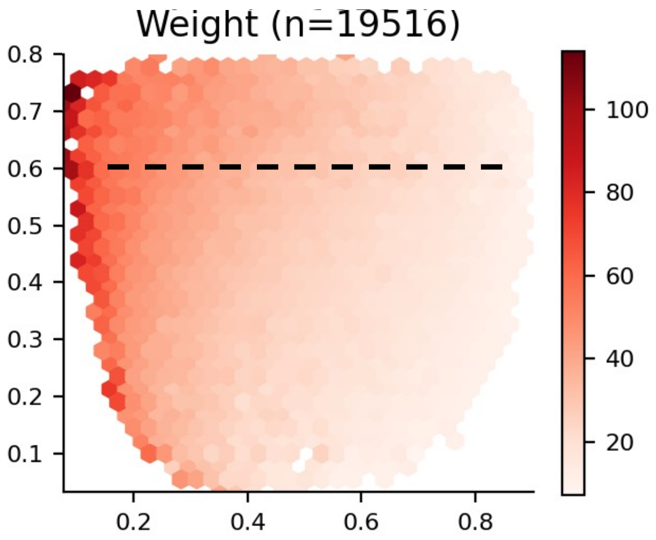


Figure 3. Latent space description: Features. The graphs represent the density distribution using hexagonal binning over the latent space.

B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Assessment and evaluation

Aligns with established characteristics of disease progression such as increase in **haematocrit levels** ↑, decrease in platelet levels ↓ and decrease in body temperature from febrile to critical phase.

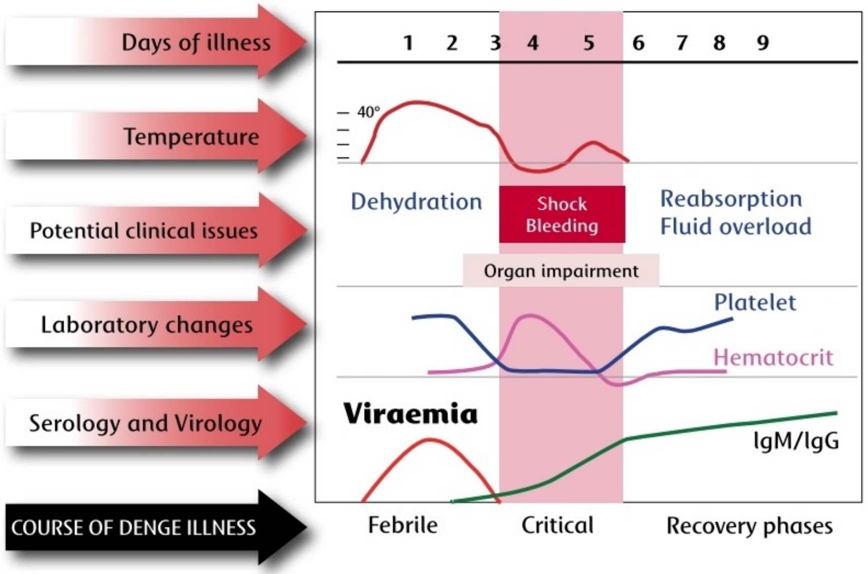


Figure 1. The course of dengue illness diagram. The figure, which has been adapted from WCL Yip, et all 1980 [28], presents phases, lab results, and associated problems.

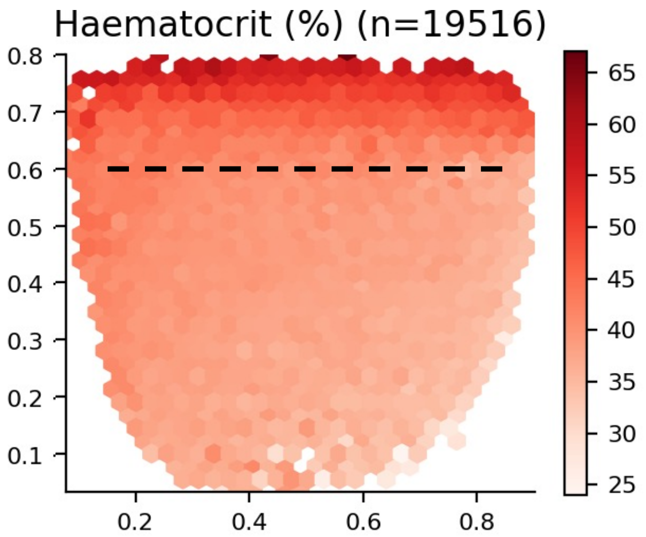


Figure 3. Latent space description: Features. The graphs represent the density distribution using hexagonal binning over the latent space.

Assessment and evaluation

Aligns with established characteristics of disease progression such as increase in haematocrit levels ↑, **decrease in platelet levels ↓** and decrease in body temperature from febrile to critical phase.

Performance
Clustering
Further analysis
Latent space
Features

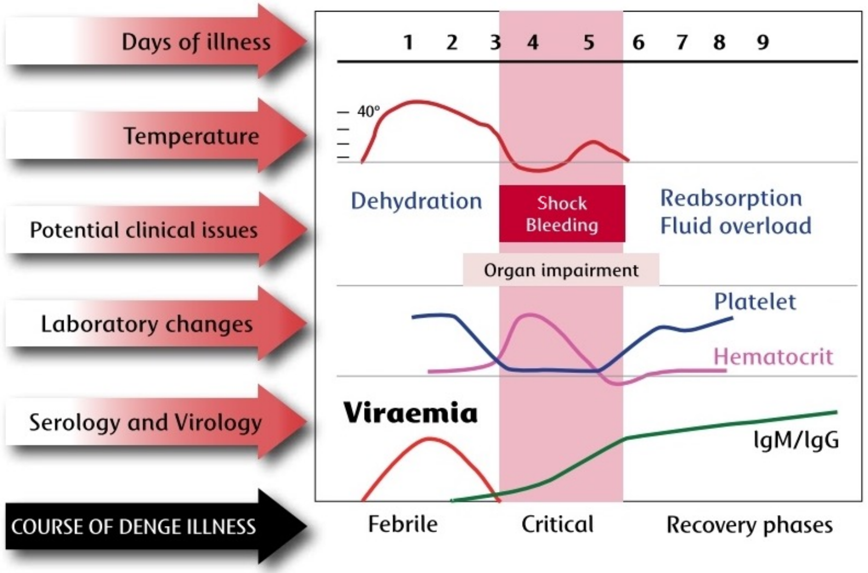


Figure 1. The course of dengue illness diagram. The figure, which has been adapted from WCL Yip, et all 1980 [28], presents phases, lab results, and associated problems.

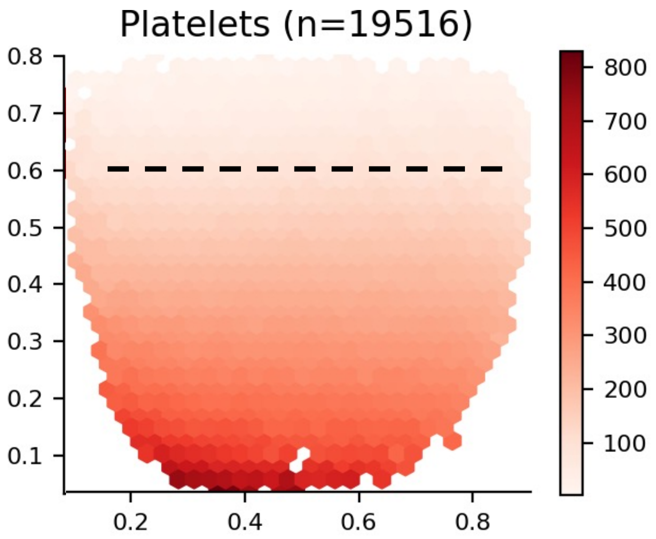


Figure 3. Latent space description: Features. The graphs represent the density distribution using hexagonal binning over the latent space.

Assessment and evaluation

Aligns with established characteristics of disease progression such as increase in haematocrit levels ↑, decrease in platelet levels ↓ and **decrease** in body temperature from **febrile to critical** phase.

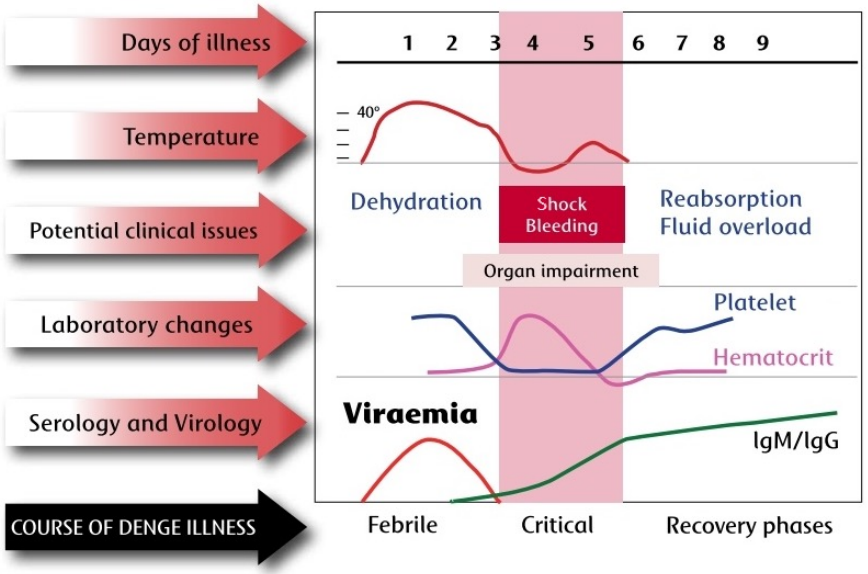


Figure 1. The course of dengue illness diagram. The figure, which has been adapted from WCL Yip, et all 1980 [28], presents phases, lab results, and associated problems.

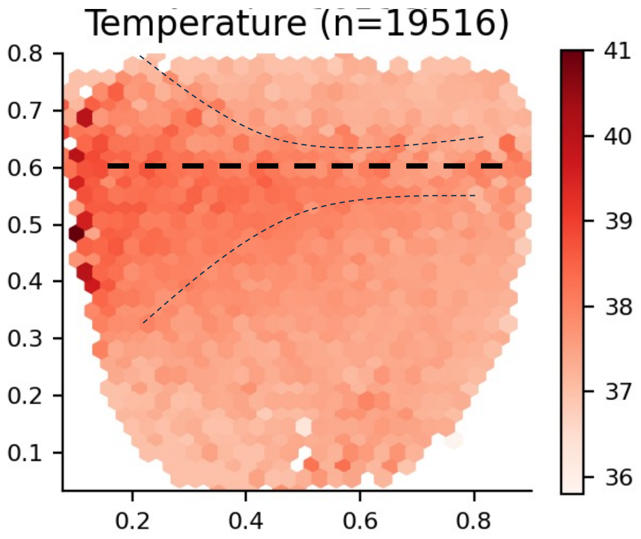
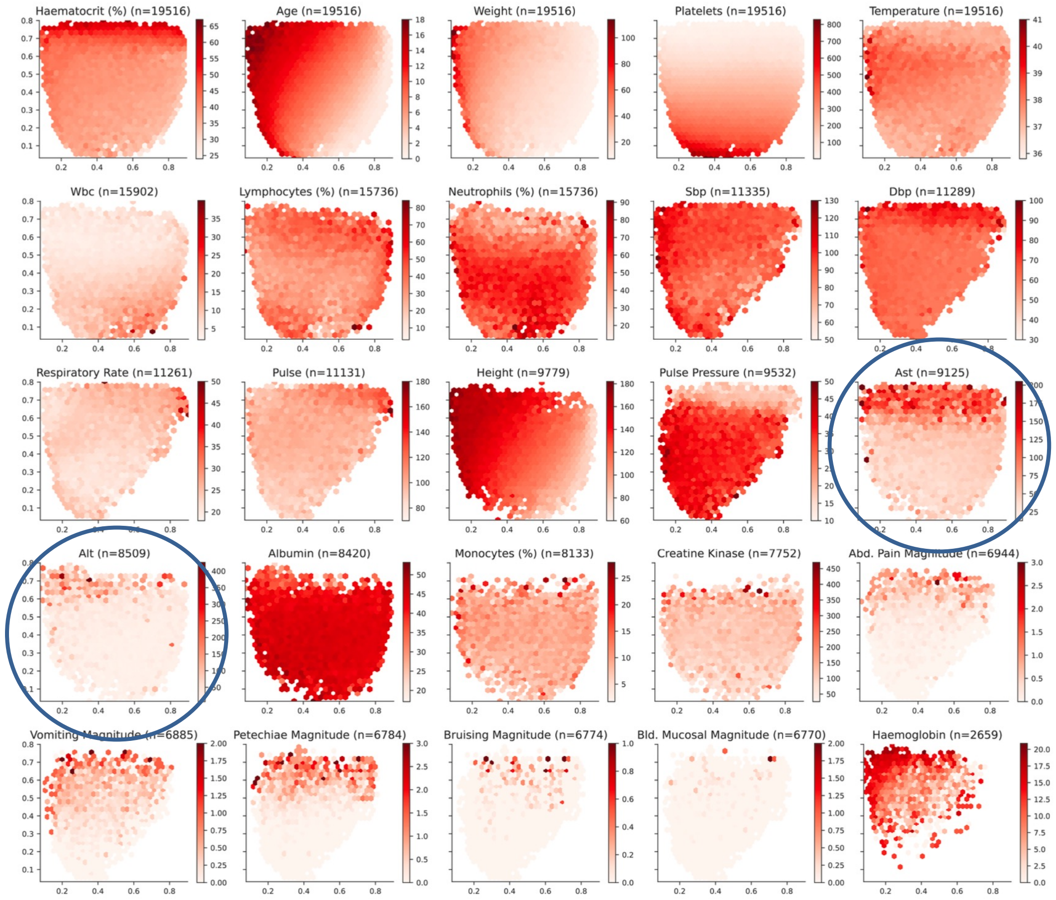


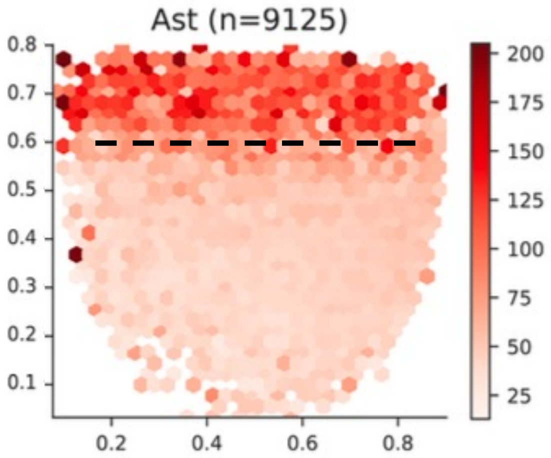
Figure 3. Latent space description: Features. The graphs represent the density distribution using hexagonal binning over the latent space.

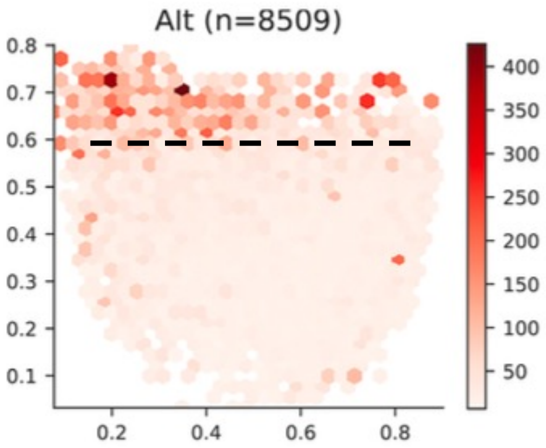
Assessment and evaluation



Appendix A1

Latent space description: Biomarkers. The graphs represent the density distribution using hexagonal binning over the latent space for the most frequent features in the dataset. The title includes the name and the number of daily profiles. The value on each hexagonal bin represents the mean value of all the daily profiles that have been projected on that bin. The magnitude values are subject to the clinicians interpretation and have been included for reference.





B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Assessment and evaluation

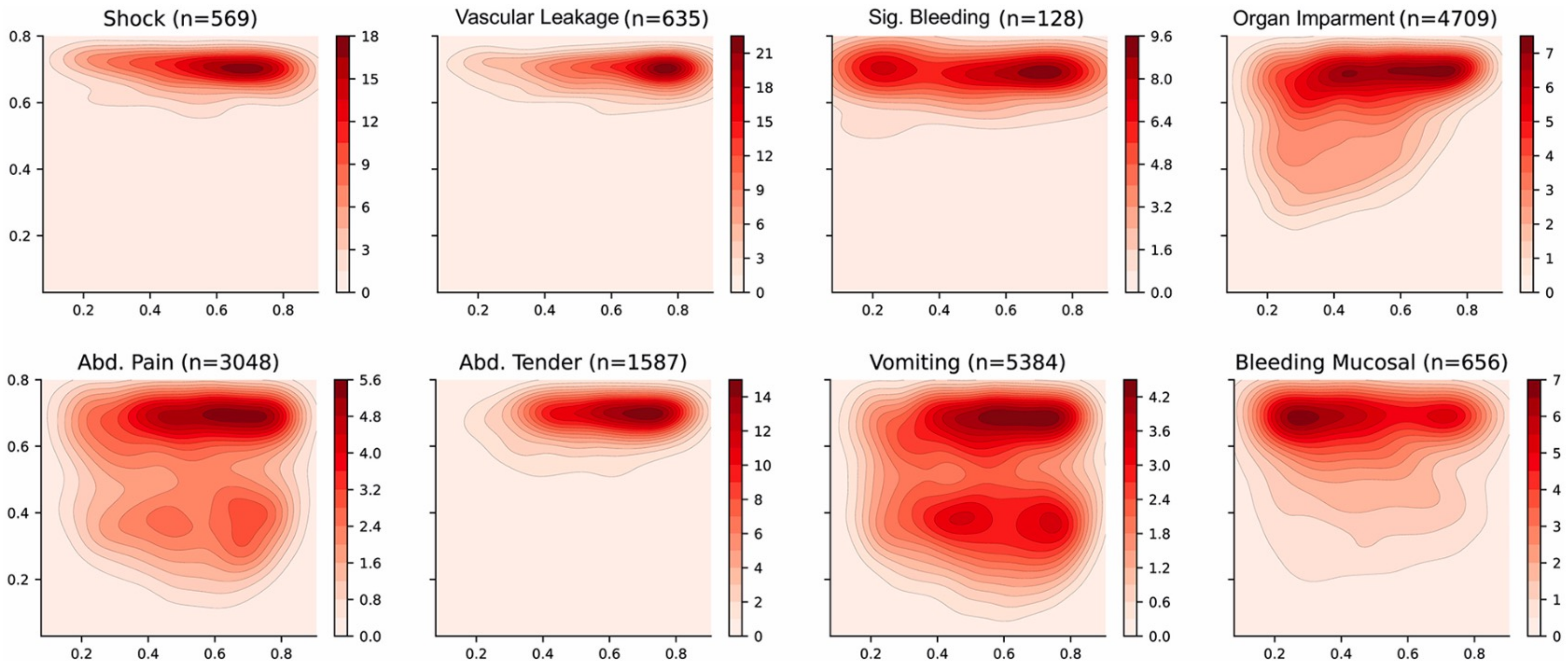


FIGURE 5

Latent space description: Phenotypes. The graphs represent the density distribution using contour lines estimated using a Gaussian kernel over the latent space. The title includes the phenotype and the number of patients in which it occurs. The definition of the compound categories (vascular leakage, significant bleeding and organ impairment) are defined in Table 2. A detailed set of graphs describing other phenotypes has been included in Figure A2 (Appendix).

World Health Organization – *Dengue: guidelines for diagnosis, treatment, prevention and control. – WHO Bulletin* (2009)
B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Assessment and evaluation

Severe dengue

Severe plasma leakage…
- leading to shock (DSS)
- fluid accumulation + distress

Severe bleeding
Severe organ involvement
- Liver: AST or ALT >= 100
- Impaired consciousness (CNS)
- Heart and other organs

Warning signs

- Abdominal pain or tenderness
- Persistent vomiting
- Fluid accumulation
- Mucosal bleeding
- Lethargy, restlessness
- Liver enlargement > 2cm
- Haematocrit ↑ & Platelets ↓

Probable dengue

Live in/travel to endemic area
Fever and 2 of:
- Nausea and/or vomiting
- Rash
- Aches and pains
- Tourniquet test positive
- Leucopenia
- Any warning sign

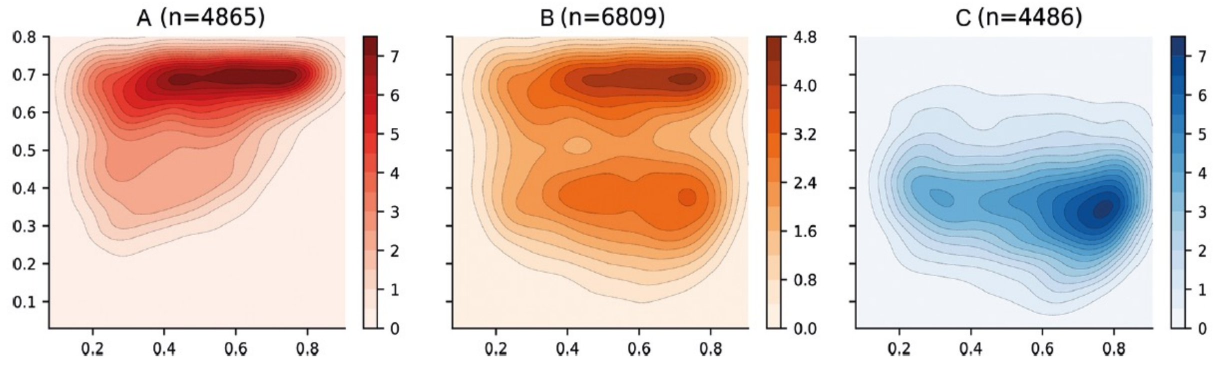


FIGURE 6
Latent space description: Categories. The graphs represent the density distribution over the latent space for three categories (A, B and C) estimated using a Gaussian kernel. These categories are defined as a compendium of various phenotypes (see Table 2) and therefore the value on each bin (pixel) described in the colorbar represents the estimated density on that bin for which one or more of the conditions associated to the category occurs.

Performance
Clustering
Further analysis
Latent space
Features
Phenotypes
Categories

World Health Organization – *Dengue: guidelines for diagnosis, treatment, prevention and control. – WHO Bulletin* (2009)
B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Assessment and evaluation

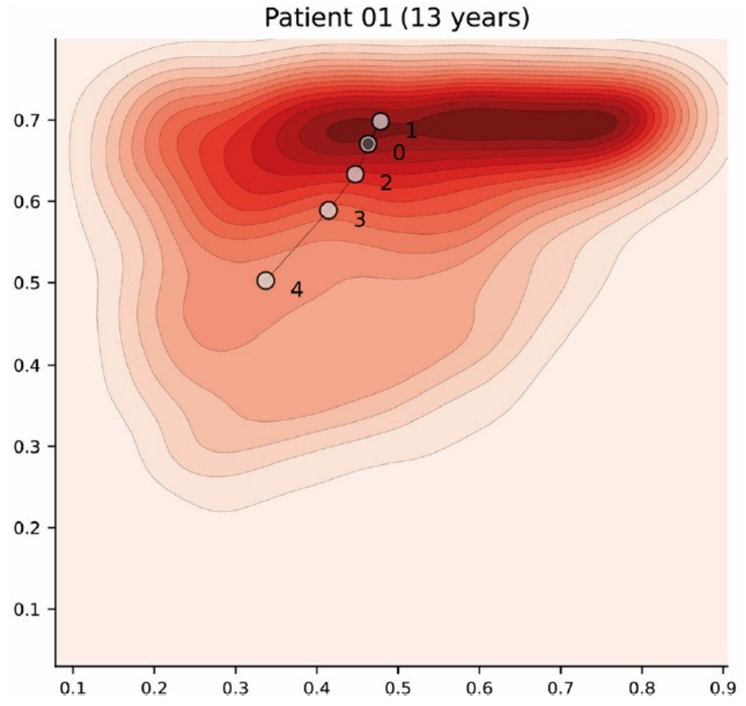


FIGURE 7
Latent space description: Trajectories. The graph represents the trajectory of a patient over the latent space using the density distribution for *Category A*, which associated with severe disease, as a background reference. Each marker represents a daily profile where the number indicates the day from admission. Filled markers indicate days in which the patient suffered an episode of shock. Further examples have been included in Figure A3 (Appendix).

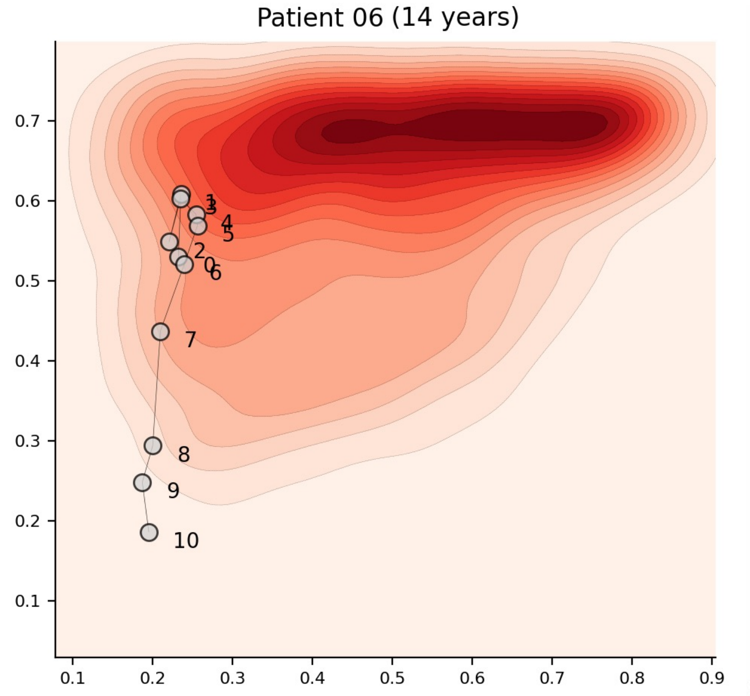
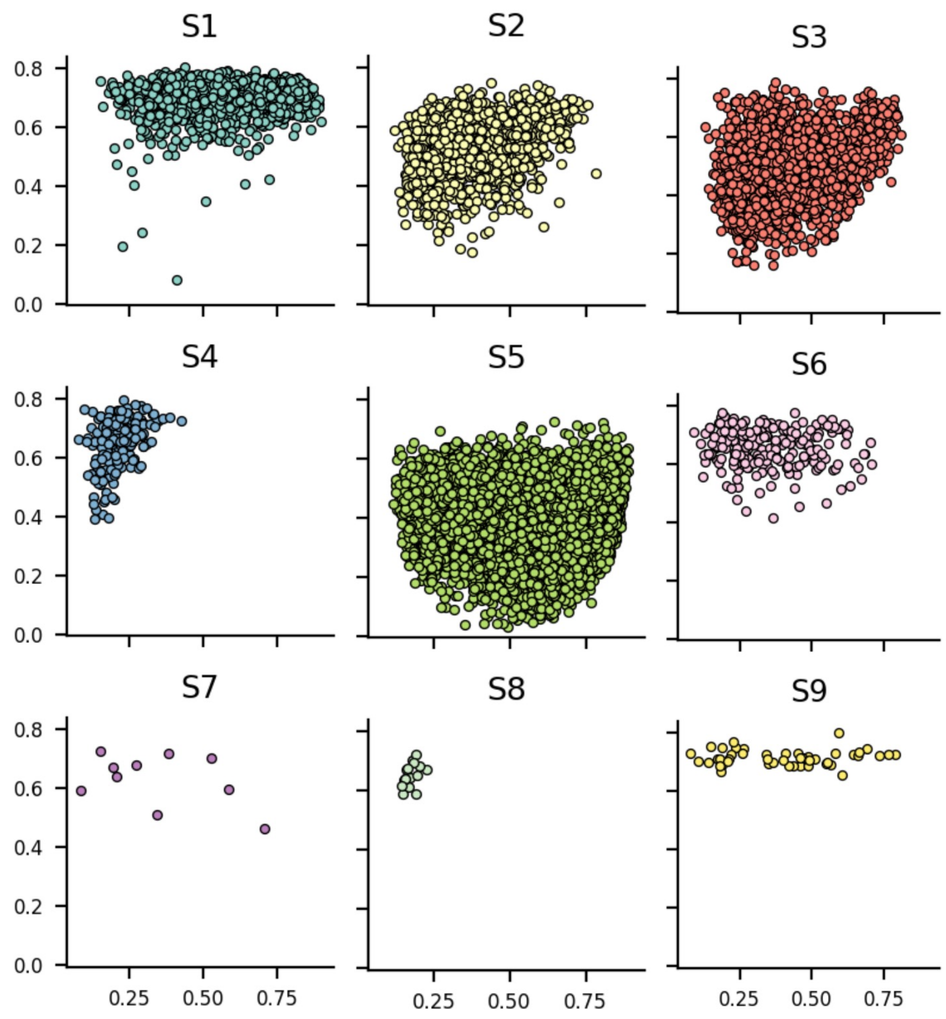


FIGURE 7
Latent space description: Trajectories. The graph represents the trajectory of a patient over the latent space using the density distribution for *Category A*, which associated with severe disease, as a background reference. Each marker represents a daily profile where the number indicates the day from admission. Filled markers indicate days in which the patient suffered an episode of shock. Further examples have been included in Figure A3 (Appendix).

Performance
Clustering
Further analysis
Latent space
Features
Phenotypes
Categories
Trajectories

Imperial College London

Assessment and evaluation



Description of studies

S1: Inpatient-based prospective observational descriptive study of clinical features of **DSS** in children and comparison of different fluid solutions for initial resuscitation.

S2: Inpatient-based prospective observational descriptive study examining prognostic factors during the **febrile phase** of non-severe dengue in children.

S4: A pilot study to investigate the effects of short course oral corticosteroid therapy in **early dengue** infection in Vietnamese patients.

S6: Laboratory diagnosis and prognosis of **severe dengue**.

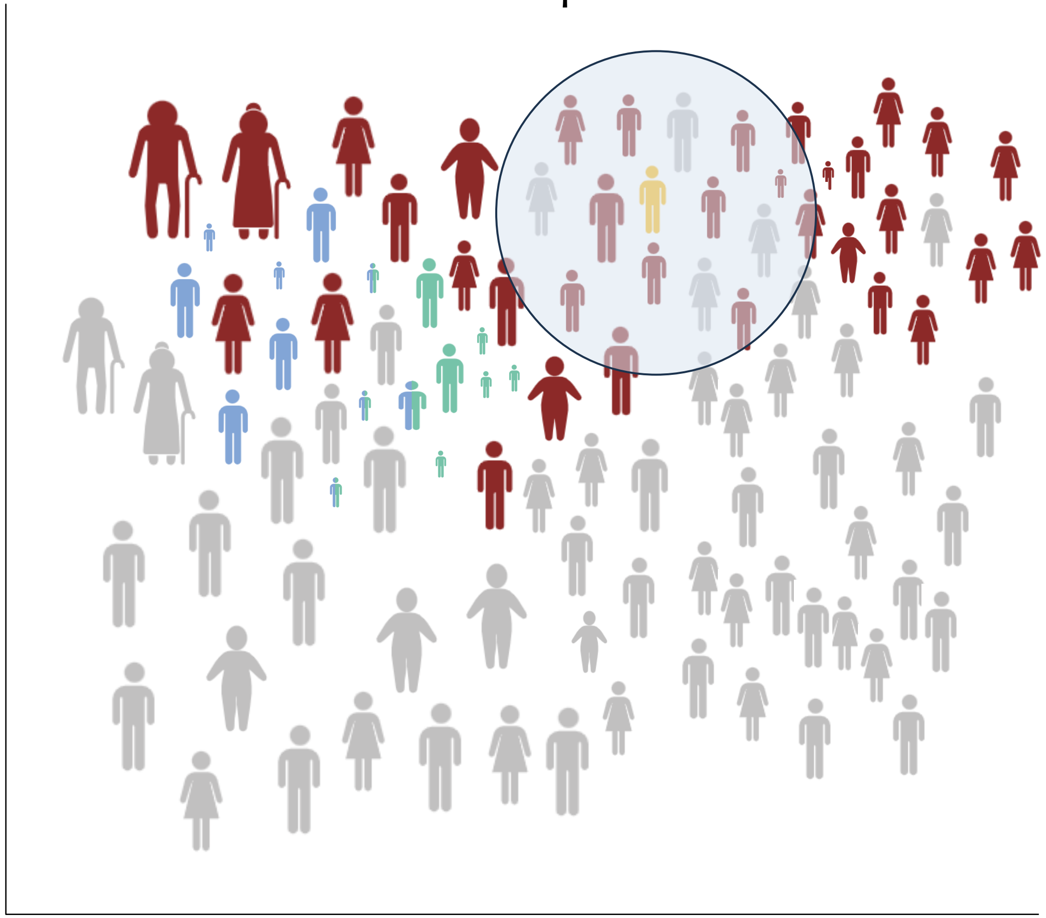
S8: A matched cohort study to characterise the clinical manifestations of dengue in **pregnancy** and investigate the spectrum of adverse maternal and foetal outcomes.

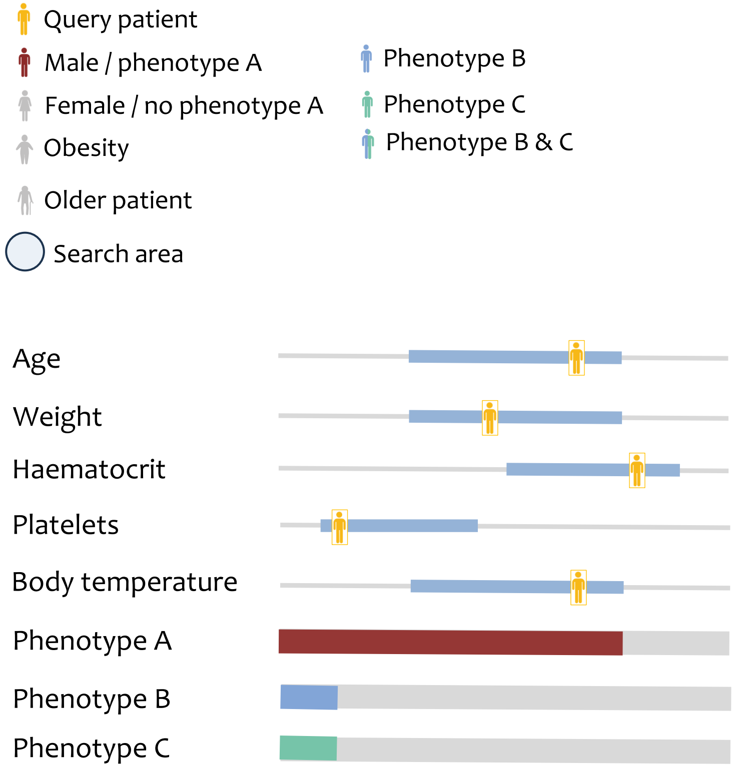
S9: Innovative biomedical engineering and computational science to improve the management of **critical illness** in resource- limited settings.

B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

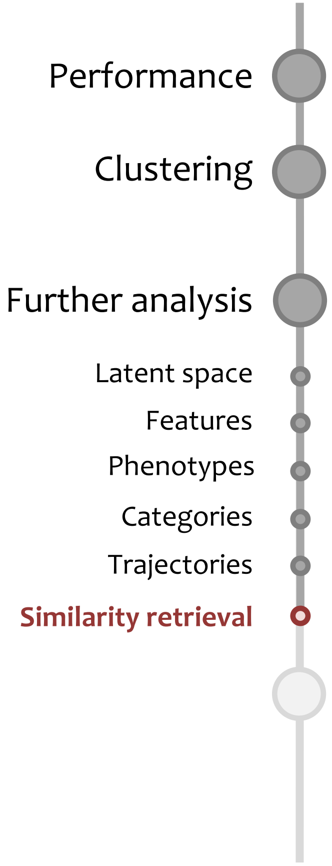
Assessment and evaluation

Latent space





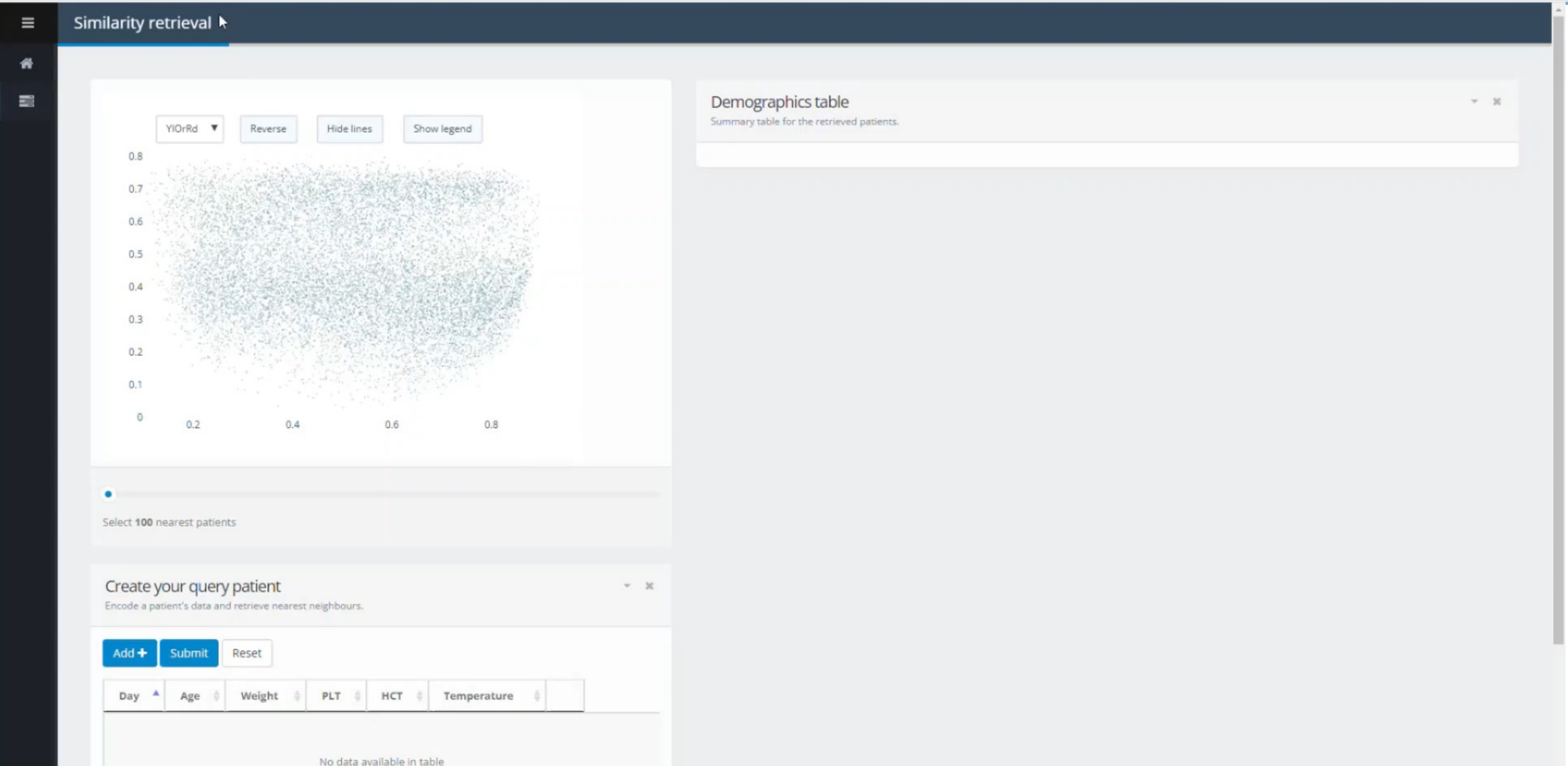
- For any feature, phenotype or category.
- If value not available, it will just not be counted.



B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Imperial College London

CDSS prototype for Dengue



Imperial College London

QUESTIONS

?

Bernard Hernandez
b.hernandez-perez@imperial.ac.uk
Department of Medicine
Imperial College London
https://bahp.github.io/portfolio-academic/

14th of December 2023

camo
centre for antimicrobial optimisation

epic impoc
enhanced, personalized and integrated care for infection management at point of care

AREAS OF RESEARCH FOR CDSS

Areas of research to leverage healthcare through clinical decision support systems

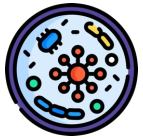
Management to facilitate access to records, laboratory results, and communication.

Facilitate **data collection** to enhance decisions, send **reminders** or clinical **alerts**

Patient engagement to improve understanding and shape future patient behaviour

Assess or predict complications using machine learning methods on tabular patient data

Review of **past similar cases** to improve decisions matching clinicians' reasoning process



AMR surveillance to strengthen knowledge for action in support of strategies to improve guidelines

Support **medication** from allergy checking, and therapy selection to adjusting antimicrobial doses

Using **NLP** to extract medical information from free-text clinical notes

Image processing to extract useful information from medical imaging (X-ray, CT, MRI)

+ more!

Imperial College London

SUPERVISED LEARNING

6, 5, 4

Infection

>500.000 daily profiles
2.7% prevalence

6 biomarkers
CRP, WBC, BIL, CRE, ALT, ALP

SVM
AUC ROC 0.85 (95% CI:0.84 - 0.86)
SENS 0.75 | SPEC 0.91

104 patients
35% prevalence at 72h
42% had microbiological tests

AUC ROC 0.84 (95% CI:0.76 - 0.91)
SENS 0.89 | SPEC 0.63 at 0.81

3

SARS-CoV-2

1186 patients
65% prevalence
14% had microbiological tests

21 biomarkers
Full Blood Count (FBC)

SVM
AUC ROC 0.91 (95% CI:0.76 - 0.91)
SENS 0.80 | SPEC 0.89

54 patients
52% prevalence

AUC ROC (0.96% CI: 0.90 – 1.00)
SENS 0.75 | SPEC 0.90

2

Dengue

8100 patients
27.7% prevalence

Age, Gender, Day + 4 biomarkers
HCT, PLT, WBC, LYMPH

XGB
AUC ROC 0.86 (95% CI:0.84 - 0.86)
SENS 0.92 | SPEC 0.56
PPV 0.73 | NPV 0.84

+ seasonality!

1

Dengue Shock Syndrome

4131 patients
5.4% prevalence

Age, Gender, Weight, Day + 2 biomarkers first 48h
HCT, PLT

ANN
AUC ROC 0.83 (95% CI:0.76 - 0.85)
SENS 0.66 | SPEC 0.84
PPV 0.18 | NPV 0.98

Bloodstream Infection

using temporal dynamics trough LSTM

1. Damien et al – *The diagnosis of dengue in patients presenting with acute febrile illness using supervised machine learning and impact of seasonality* – Frontiers in Digital Health (2022)
2. Damien et al – *Applied machine learning for the risk-stratification and clinical decision support of hospitalised patients with dengue in Vietnam* – PLOS Digital Health (2022)
3. TM Rawson et al – *Supervised machine learning to support the diagnosis of bacterial infection in the context of COVID-19* – JAC Antimicrobial Resistance – BMC Medicine (2021)
4. B Hernandez et al – *Data-driven web based intelligent CDSS for infection management at the point of care* – PhD Thesis Imperial College London (2019)
5. TM Rawson et al – *Supervised machine learning for the prediction of infection on admission to hospital: a prospective observational cohort study* – JAC (2018)
6. B Hernandez et al – *Supervised learning for infection risk inference using pathology data* – BMC Medical Informatics and Decision Making (2017)

Results

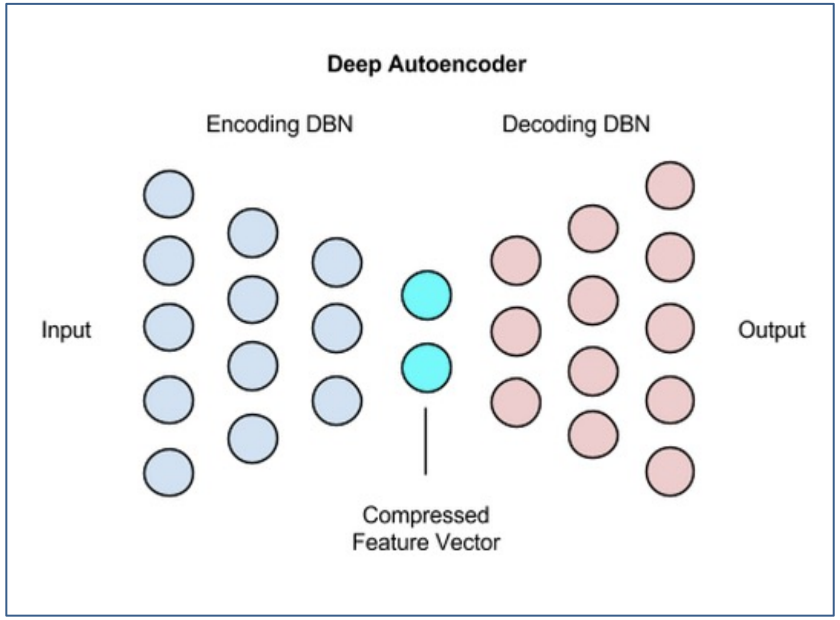
The **input features** selected are:

- Age (years)
- Weight (Kg)
- Body Temperature (°C)
- Platelets (k/µ L)
- Haematocrit (%)

- Features were consistently recorded in all clinical studies.
- Provide information on the course of dengue illness (WHO).
- Feasibility of prospective collection

Algorithm	Type	Metrics	Comments
PCA	Parametric	Good distance metrics. Density metrics are decent but inferior to SOM/AE.	Performance of PCA is likely to decrease as dimensionality increases due to its linear nature.
t-SNE	Non-parametric	Performance highly dependent on hyperparameters	High computationally expensive to create embeddings for **unseen data.** Can create completely separated clusters.
SOM	Non-parametric	Poor distance preservation. Good density metrics.	Limitations imposed by the **discrete space** limit its usability for similarity retrieval.
AE	Parametric	Good performance for both distance and density metrics.	Quite a flexible approach, possible to use with time-series signals or even images.

[1] UMAP might be also a promising algorithm.



Assessment and evaluation

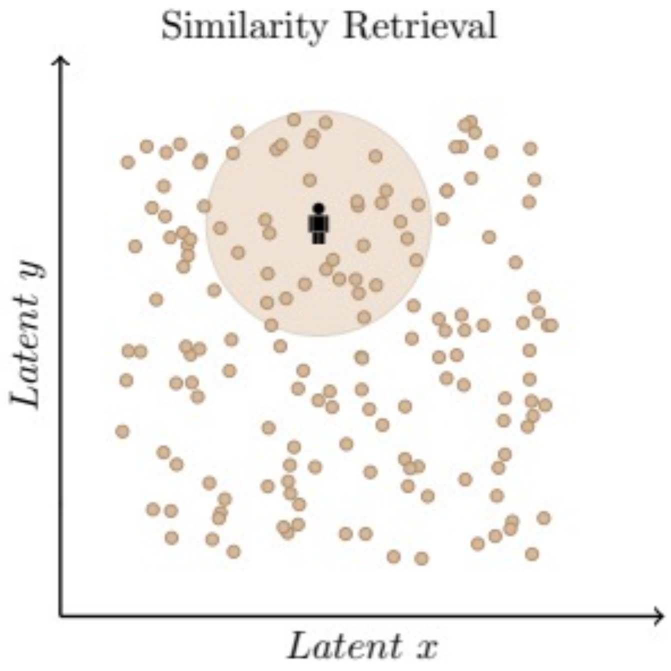
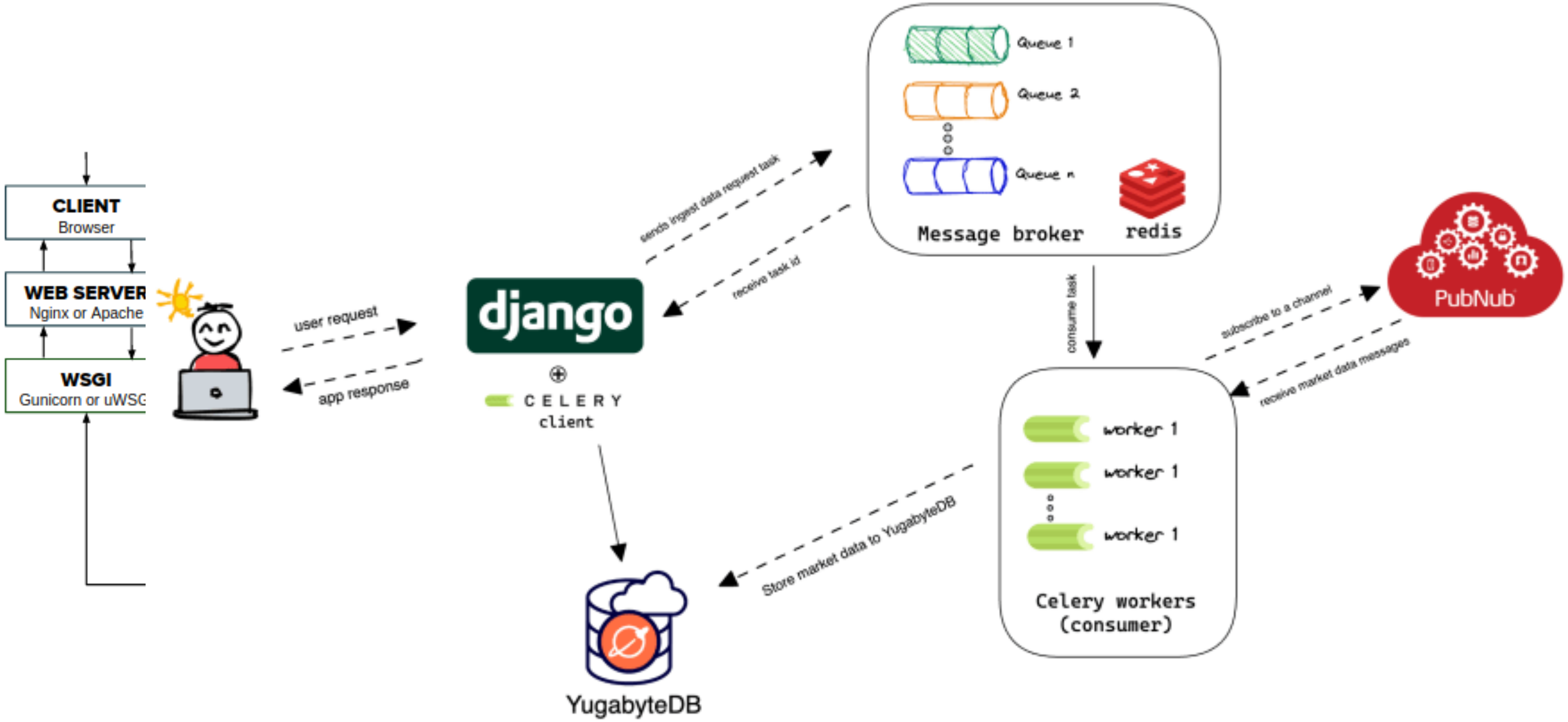


Figure 8. Latent space description: Similarity retrieval. The diagram represents a query patient and the area of interest for which patients, which are deemed to be similar, should be retrieved.

B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)

Assessment and evaluation



request-response lifecycle in Django

- Performance
- Clustering
- Further analysis
  - Latent space
  - Features
  - Phenotypes
  - Categories
  - Trajectories
  - Similarity retrieval
- Deployment

B. Hernandez et al – *Learning meaningful latent space representations for patient risk stratification: Dengue and other acute febrile illness – Frontier Digital Health* (2023)