Towards the development of a continuous model for the assessment of bacteremia, bloodstream infection and sepsis

Óscar Escudero-Arnanz
oscar.escudero@urjc.es
Department of Signal Theory and Communications, Telematics and Computing Systems
Rey Juan Carlos University
https://github.com/oscarescuderoarnanz

Joint work with Bernard Hernandez, Cristina Soguero-Ruiz, Inmaculada Mora-Jiménez, and A. G. Marques

June 21st, 2024

Academic journey

Rey Juan Carlos University (URJC), Madrid, Spain

- B.Sc. in Telecommunications
- M.Sc. in Telecommunications
- Ph.D. in Artificial Intelligence and Data Science

Università Campus Bio-Medico (UCBM), Roma, Italy

- Visiting researcher (PhD)

Imperial College London (ICL), London, United Kingdom

- Visiting researcher (PhD)

URJC

Academic journey

IMPERIAL

Rey Juan Carlos University (URJC), Madrid, Spain. **Ph.D. in Artificial Intelligence and Data Science**

- Electronic Health Record
- Irregular Multivariate Time Series
- Distance measures and kernel methods
- Multimodal Architectures
- Interpretable Recurrent Neural Networks
- Interpretable Spatio-Temporal Graph Neural Networks

→ Clinical Task

[1] Escudero-Arnanz Ó, Mora-Jiménez I, Martínez-Agüero S, Álvarez-Rodríguez J, Soguero-Ruíz C. Temporal Feature Selection for Characterizing Antimicrobial Multidrug Resistance in the Intensive Care Unit. InAAI4H@ ECAI 2020 Sep (pp. 54-59)

[2] Escudero-Arnanz, Óscar, Rodríguez-Álvarez, J., Mikalsen, K. Ø., Jenssen, R., & Soguero-Ruiz, C. On the Use of Time Series Kernel and Dimensionality Reduction to Identify the Acquisition of Antimicrobial Multidrug Resistance in the Intensive Care Unit. KDD 2021 Health Day and 2021 KDD Workshop on Applied Data Science for Healthcare.

[3] Escudero-Arnanz Ó, Mora-Jiménez I, Martínez-Agüero S, Álvarez-Rodríguez J, Soguero-Ruíz C.
Feature Selection and Tree-based Models to Predict Multidrug-Resistance. CASEIB 2020.

[4] Oscar Escudero-Arnanz, Antonio G. Marques, Cristina Soguero-Ruiz, Inmaculada Mora-Jiménez, Gregorio Robles. dtwParallel: A Python Package to Efficiently Compute Dynamic Time Warping Between Time Series. SoftwareX, 2023.

[5] Oscar Escudero-Arnanz, Antonio G. Marques, Rosa Sicilia, Cristina Soguero-Ruiz. Low-Rank Tensor Completion for Heart Failure Detection in Multivariate Time Series with Missing Data. IEEE 37th International Symposium on Computer Based Medical Systems 2024.

[6] Oscar Escudero-Arnanz, Antonio G. Marques, Inmaculada Mora-Jiménez, Joaquín Álvarez-Rodríguez, Cristina Soguero-Ruiz. Leveraging Multivariate Time Series Analysis and Machine Learning for the Characterization of Antimicrobial Resistance in the Intensive Care Unit. Engineering Applications of Artificial Intelligence. (submitted, 2024, under review)

[7] Oscar Escudero-Arnanz, Cristina Soguero-Ruiz, Inmaculada Mora-Jiménez, Joaquín Álvarez-Rodríguez, Antonio G. Marques. Explainable AI Techniques for Irregular Temporal Prediction of Antimicrobial Multidrug Resistance Acquisition in Intensive Care Unit Patients. Engineering Applications of Artificial Intelligence. (to submit on 30th in June 2024)

[8] Oscar Escudero-Arnanz, Cristina Soguero-Ruiz, Inmaculada Mora-Jiménez, Joaquín Álvarez-Rodríguez, Antonio G. Marques. Explainable Spatio-Temporal Graph Architecture for Irregular Multivariate Time Series in Inference Tasks. IEEE TSIPN. (to submit on 30th in July 2024)

Problem definition

Bacteremia

The presence of bacteria in the bloodstream, which can originate from various sources such as wounds, surgical procedures, or other infections

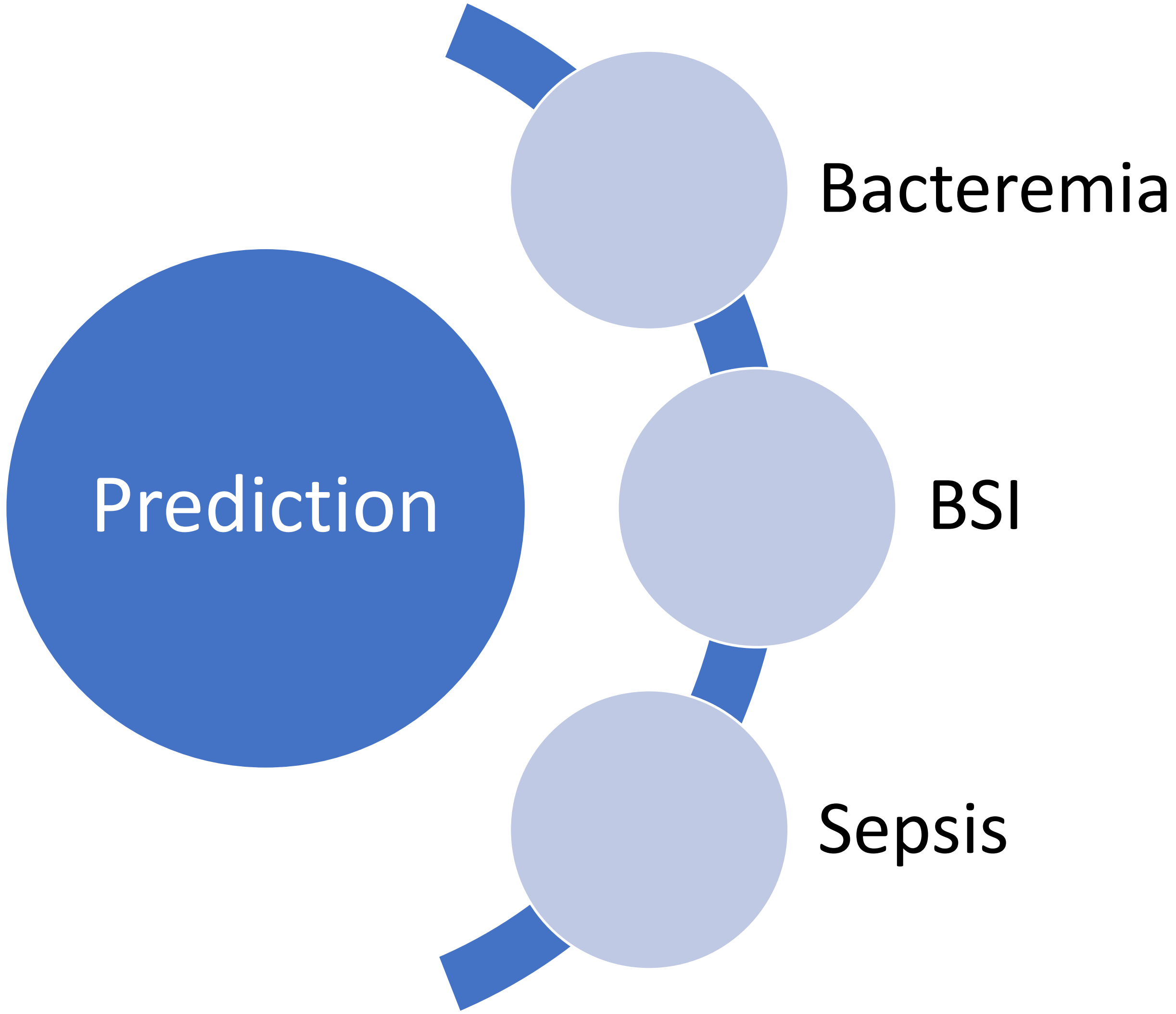
Bloodstream Infection (BSI)

Bacteremia with systemic signs of infection

Sepsis

A severe systemic response to infection, often marked by widespread inflammation, organ dysfunction, and a high risk of mortality

Objective



URJC

Databases & inclusion criteria

IMPERIAL

Data Sources

- We will use data from four public datasets…

MIMIC	AUMC
eICU	HiRID

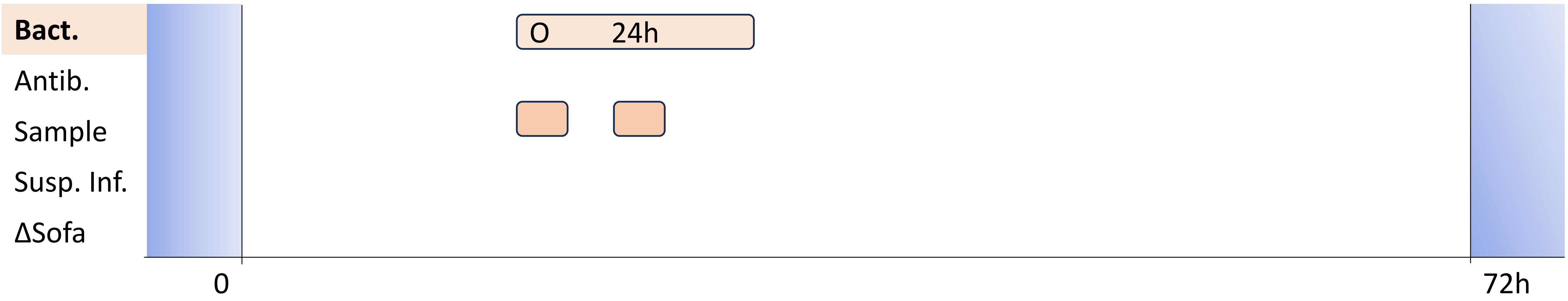
- Focusing on patients in the Intensive Care Unit (ICU).

Inclusion Criteria for Patients

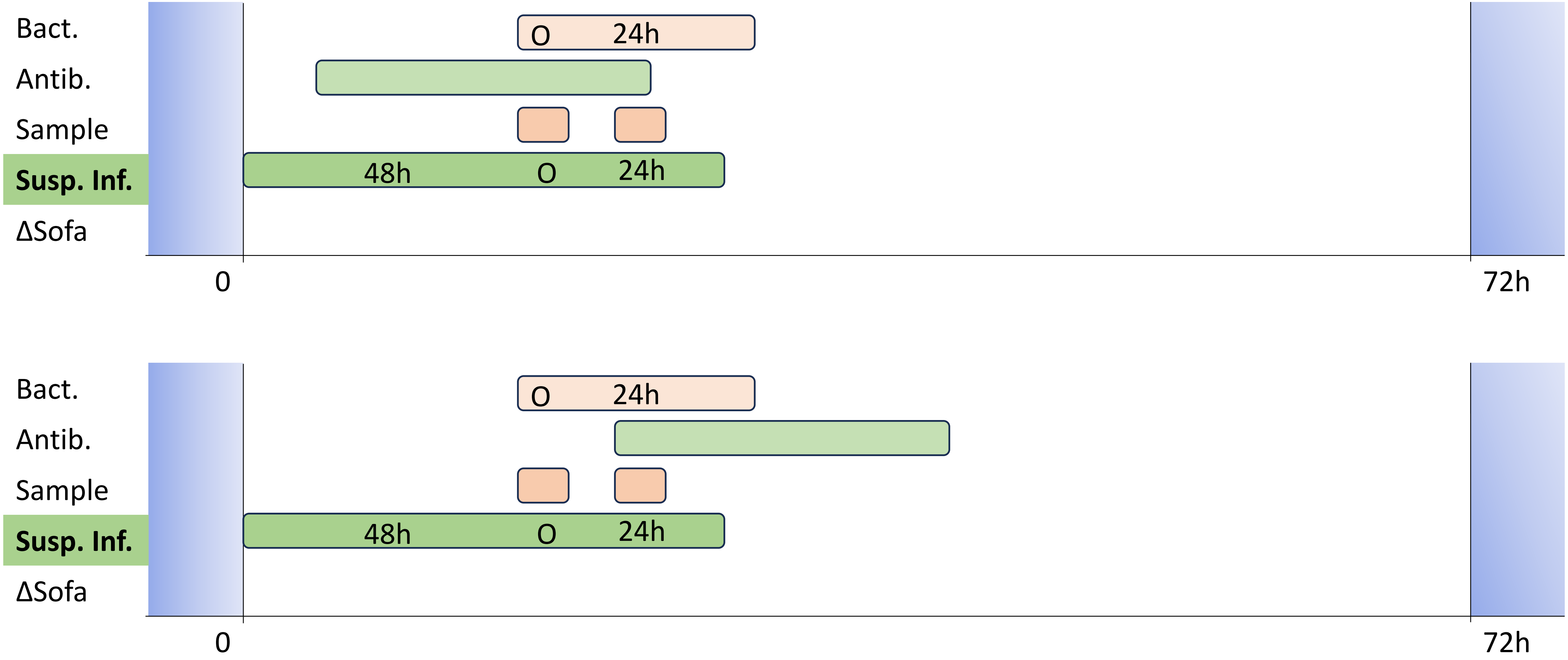
- Exclude patients with more than 85% missing values.

55	7	4
Biochemical markers	Vital signs	Statics

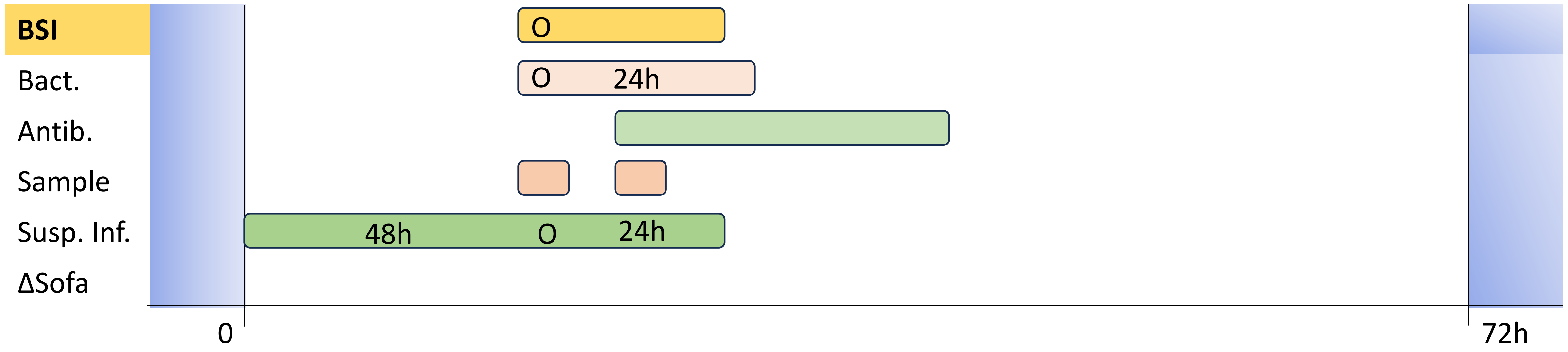
Labels definition. Bacteremia



Labels definition. Suspected Infection



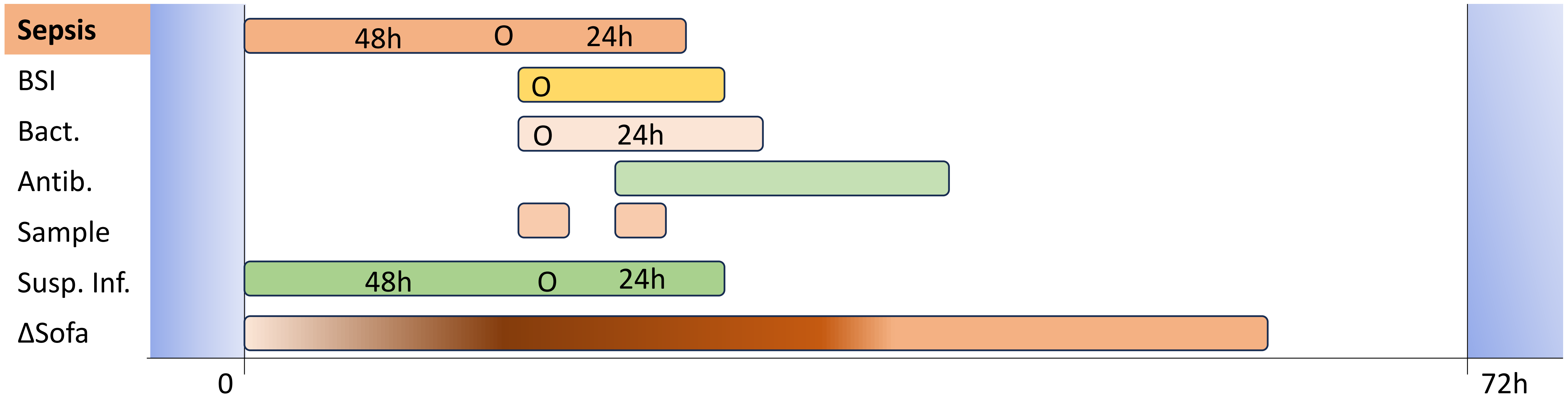
Labels definition. BSI



URJC

Labels definition. Sepsis

IMPERIAL



Pre-processing

Pre-processing

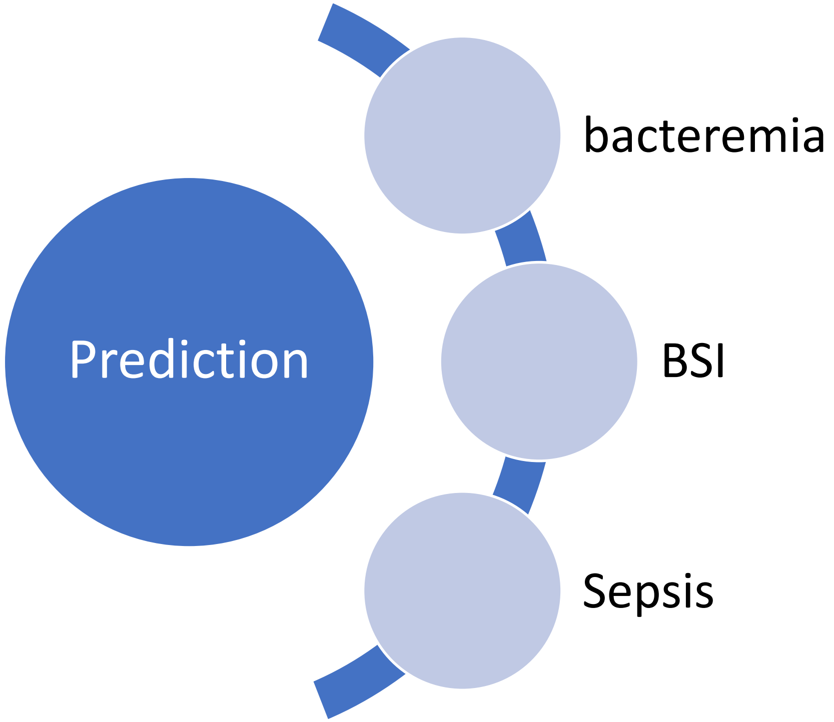
Each patient has been segmented into windows

- **Window Size:** Fixed at 6 hours
- **Look Back Period:** 6 hours
- **Look Ahead Period:** 1 hour

Prediction Goal

The aim is to predict within a 6-hour look back and 1-hour look ahead window whether the patient will develop:

- Bacteriemia
- BSI
- Sepsis



Approaches

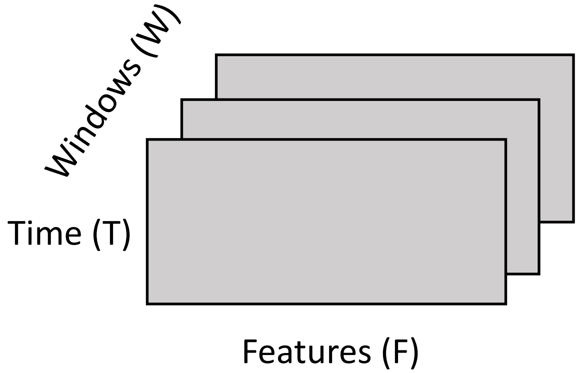
Post Preprocessing

$D_{vs}^{t} = \{(\boldsymbol{X}_p)\}_{p=1}^{P}$, with $\boldsymbol{X}_p \in \mathbb{R}^{WxFxT}$

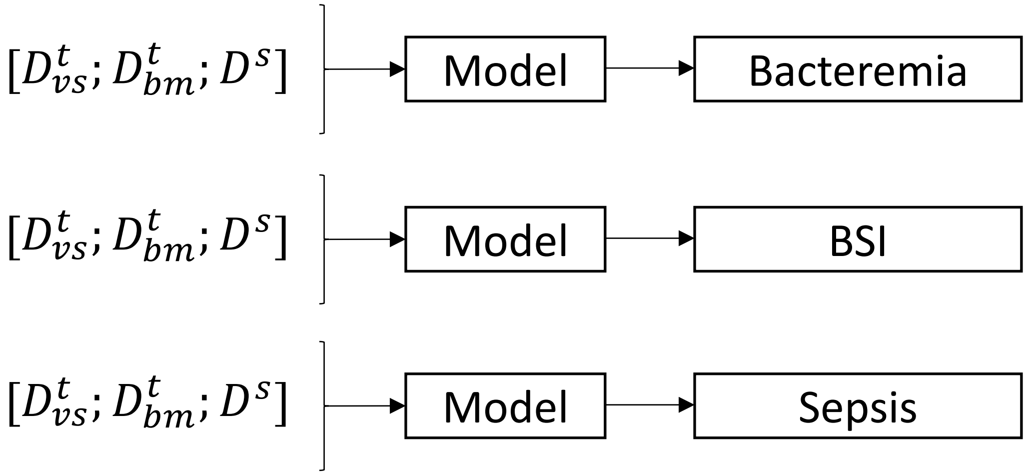
$D_{bm}^{t} = \{(\boldsymbol{X}_p)\}_{p=1}^{P}$, with $\boldsymbol{X}_p \in \mathbb{R}^{WxFxT}$

$D^{s} = \{(\boldsymbol{x}_p)\}_{p=1}^{P}$, with $\boldsymbol{x}_p \in \mathbb{R}^{WxF}$

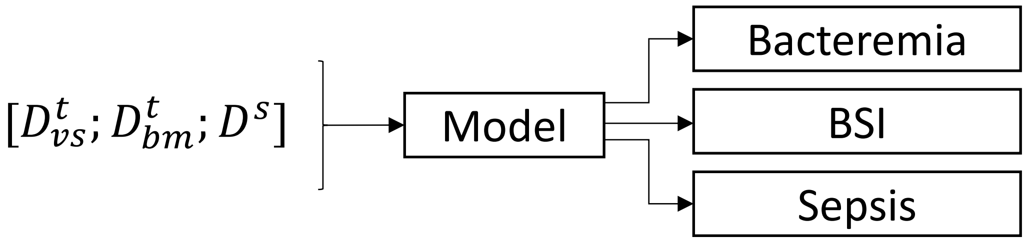
$D_{t} = \{(\boldsymbol{Y}_p)\}_{p=1}^{P}$, with $\boldsymbol{Y}_p \in \mathbb{R}^{WxLxT}, beeing\ L = \{Bac, BSI, Sepsis\}$



Approach 1



Approach 2



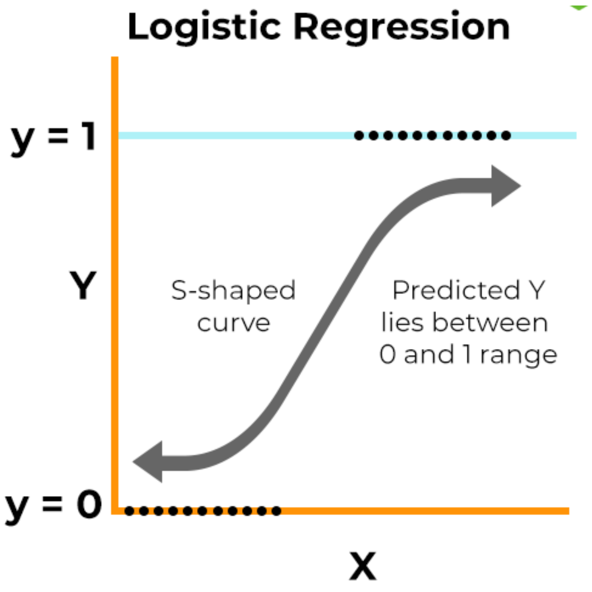
Approaches

Approach 1

$[D_{vs}^{t}; D_{bm}^{t}; D^{s}]$ → LR → Bacteremia

$[D_{vs}^{t}; D_{bm}^{t}; D^{s}]$ → LR → BSI

$[D_{vs}^{t}; D_{bm}^{t}; D^{s}]$ → LR → Sepsis



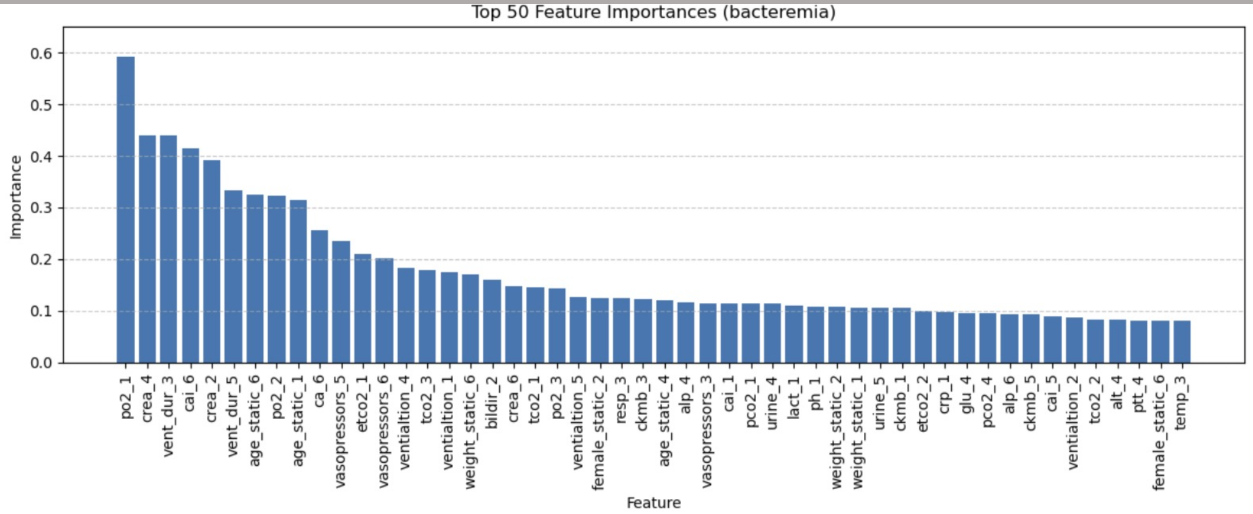
Results

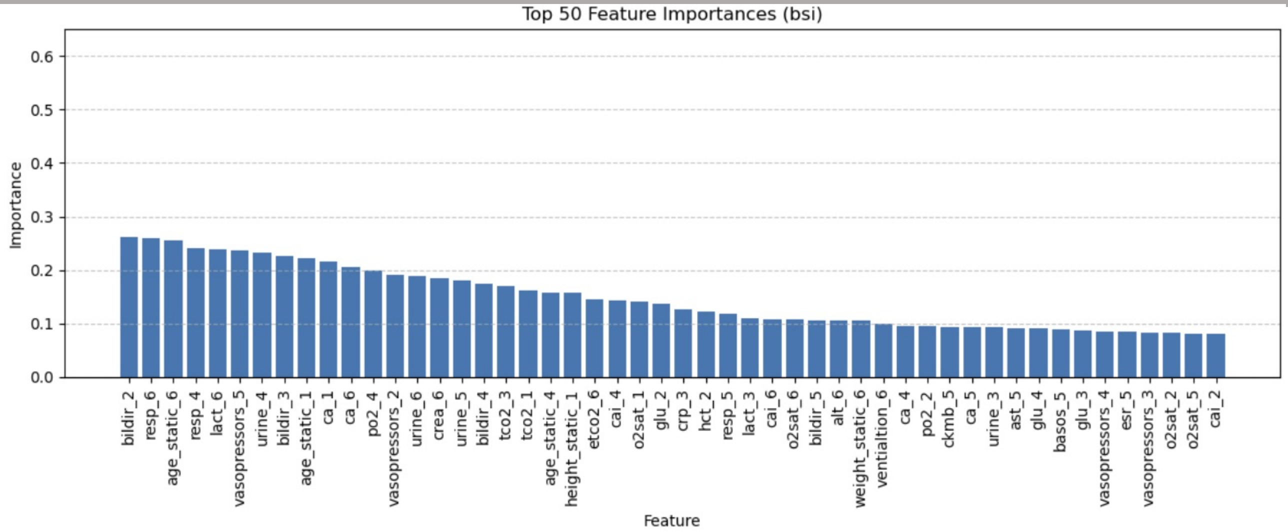
Bacteremia	BSI	Sepsis
Specificity: 74.6 +- 2.58	Specificity: 69.91 +- 2.65	Specificity: 70.15 +- 2.65
Sensitivity: 61.44 +- 6.88	Sensitivity: 48.39 +- 7.2	Sensitivity: 55.58 +- 11.45
AUC: 76.97 +- 2.5	AUC: 64.99 +- 1.37	AUC: 67.17 +- 4.22

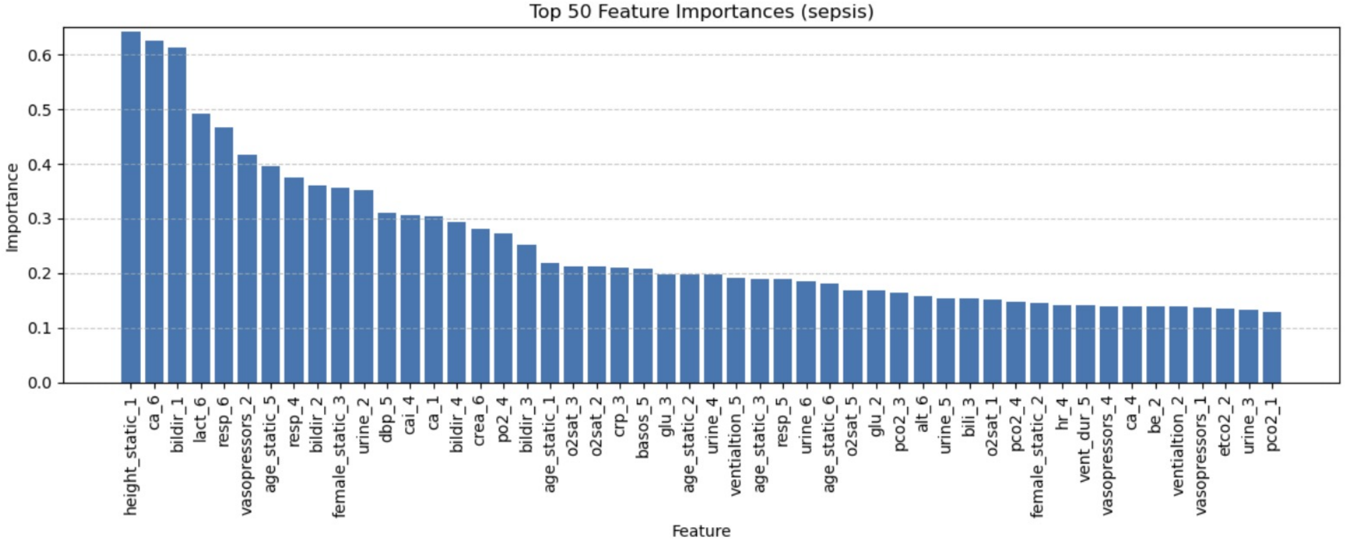
URJC

Approaches

IMPERIAL







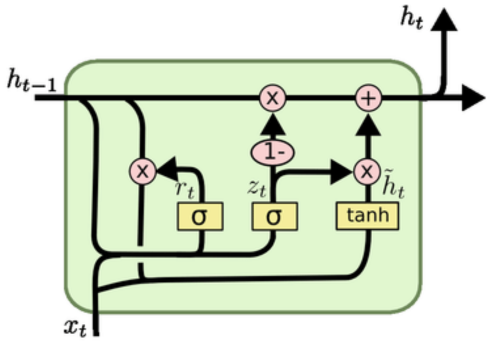
Approaches

Approach 1

$[D_{vs}^t; D_{bm}^t; D^s]$ → GRU → Bacteremia

$[D_{vs}^t; D_{bm}^t; D^s]$ → GRU → BSI

$[D_{vs}^t; D_{bm}^t; D^s]$ → GRU → Sepsis

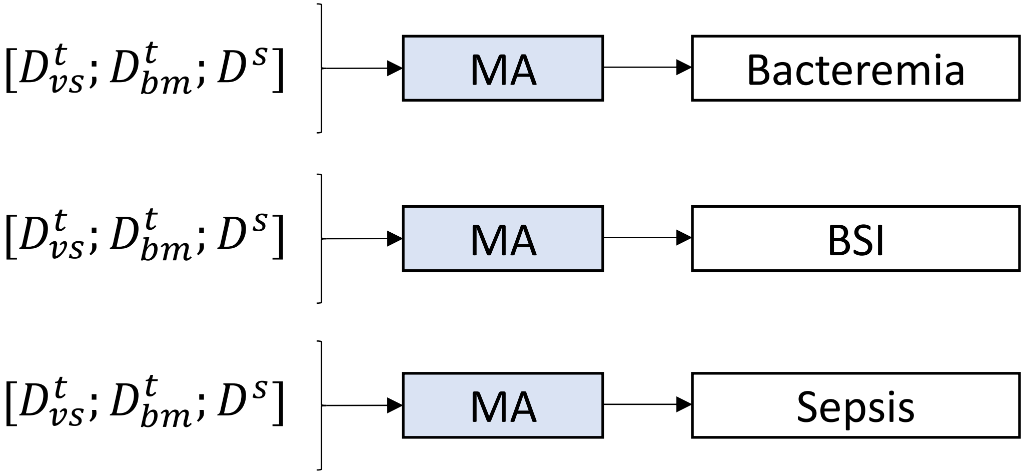


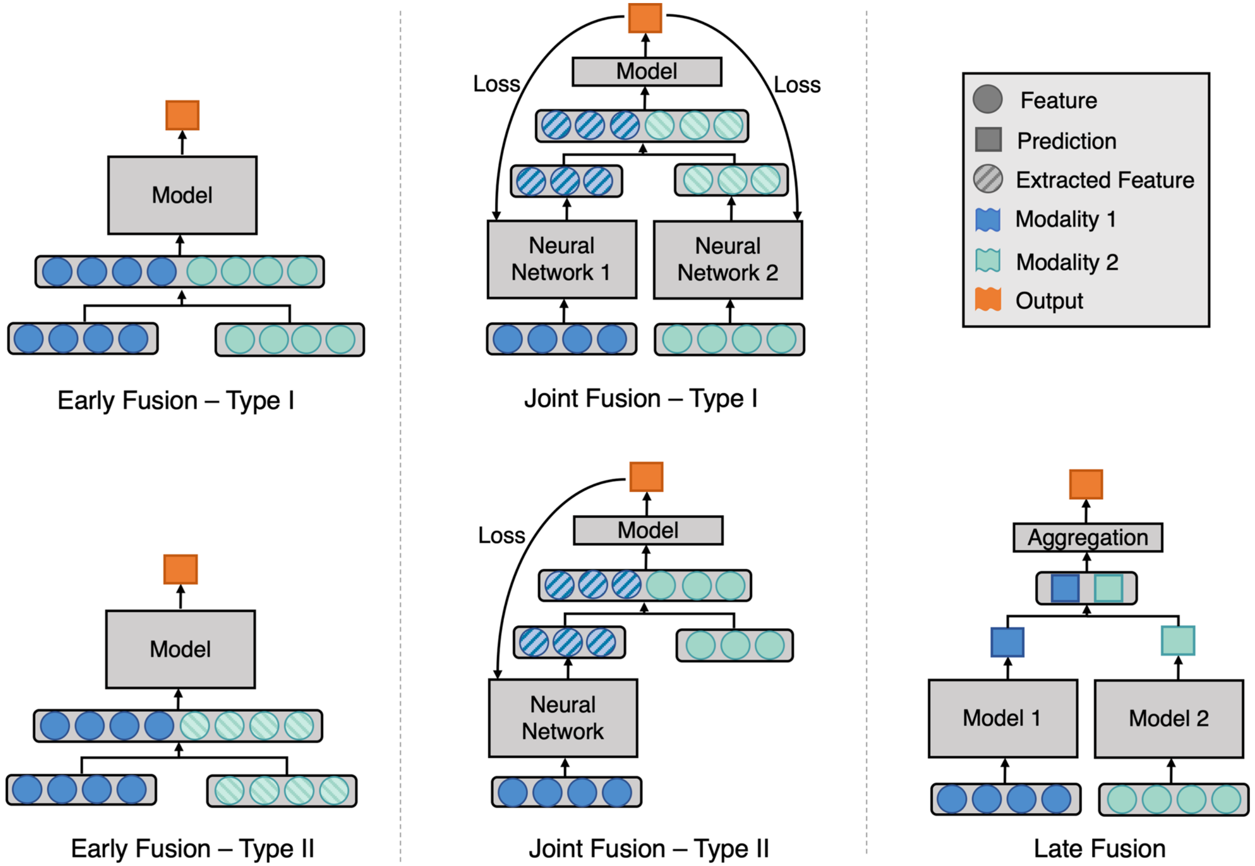
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$
$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$
$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

Results		
Bacteremia	BSI	Sepsis
Running...	Running...	Running...

Approaches

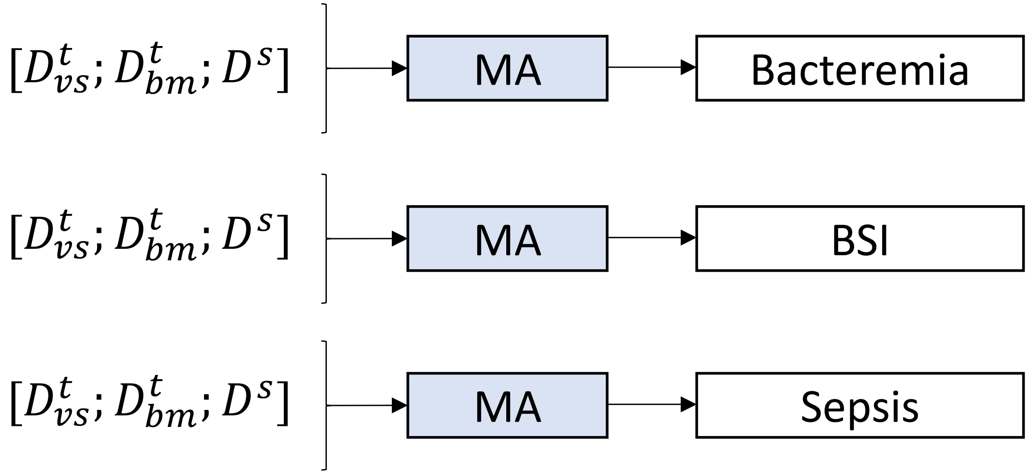
Approach 1

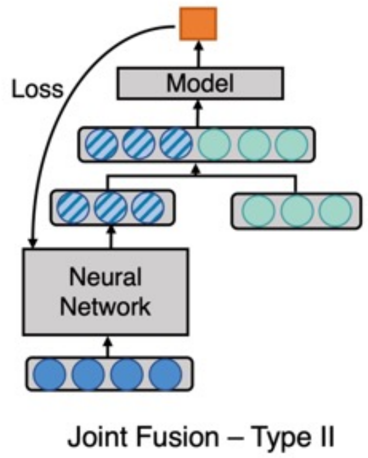


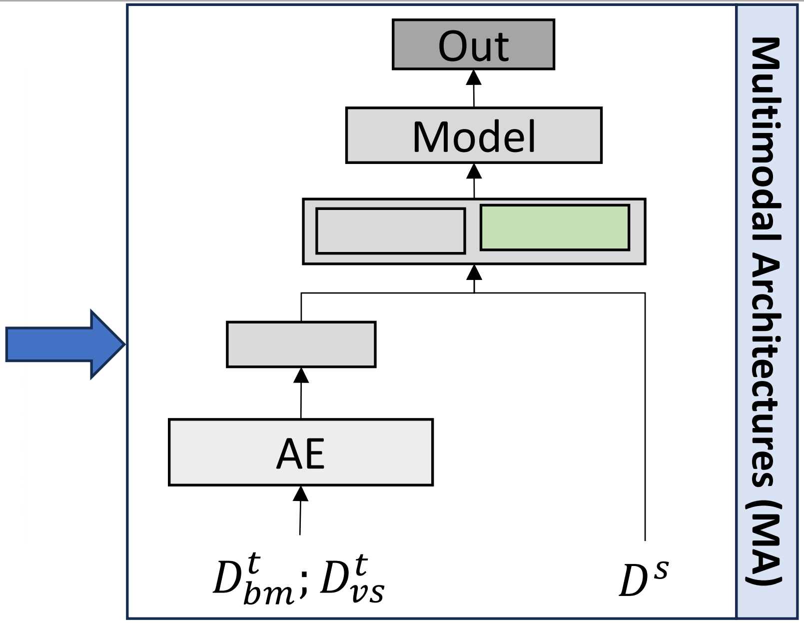


Approaches

Approach 1







Results

Bacteremia

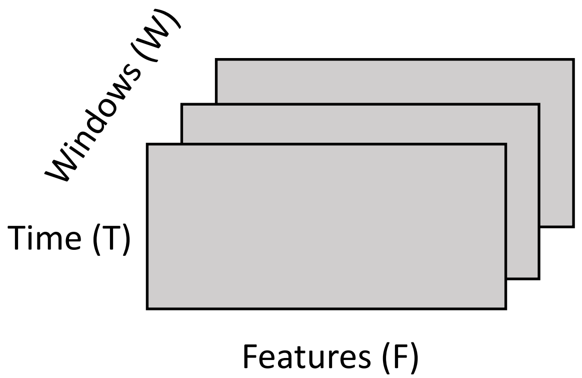
Specificity: 71.29 +- 3.27
Sensitivity: 64.26 +- 7.86
AUC: 75.67 +- 2.72

BSI

Specificity: 69.67 +- 3.1
Sensitivity: 49.49 +- 7.98
AUC: 65.3 +- 1.92

Sepsis

Specificity: 70.02 +- 3.39
Sensitivity: 56.93 +- 10.25
AUC: 68.34 +- 3.2



Post Preprocessing

$D_{vs}^{t} = \{(\boldsymbol{X}_p)\}_{p=1}^{P}$, with $\boldsymbol{X}_p \in \mathbb{R}^{WxFxT}$

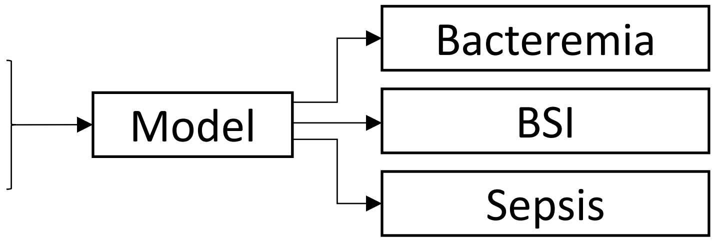
$D_{bm}^{t} = \{(\boldsymbol{X}_p)\}_{p=1}^{P}$, with $\boldsymbol{X}_p \in \mathbb{R}^{WxFxT}$

$D^{s} = \{(\boldsymbol{x}_p)\}_{p=1}^{P}$, with $\boldsymbol{x}_p \in \mathbb{R}^{WxF}$

$D_{t} = \{(\boldsymbol{Y}_p)\}_{p=1}^{P}$, with $\boldsymbol{Y}_p \in \mathbb{R}^{WxLxT}, beeing\ L = \{Bac, BSI, Sepsis\}$

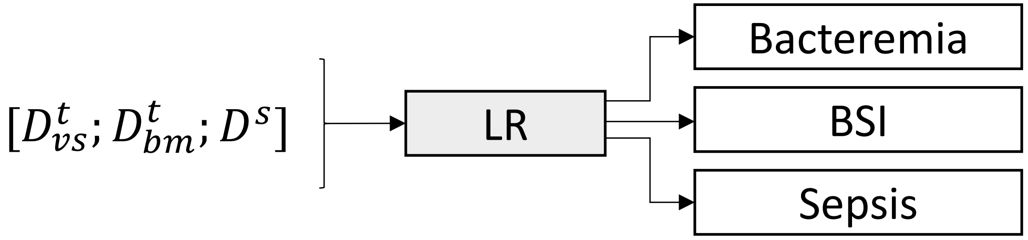
Approach 2

$[D_{vs}^{t}; D_{bm}^{t}; D^{s}]$



Approaches

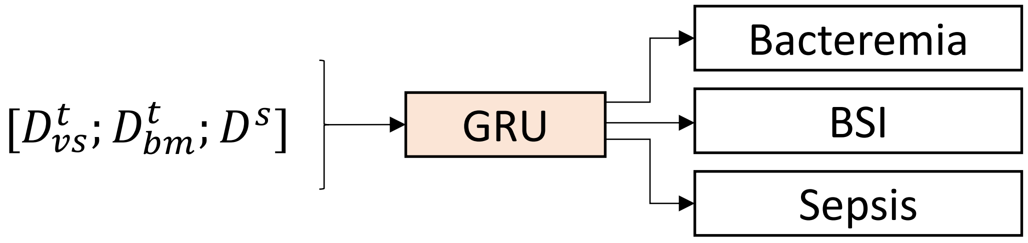
Approach 2



Results		
Bacteremia	BSI	Sepsis
Running...	Running...	Running...

Approaches

Approach 2

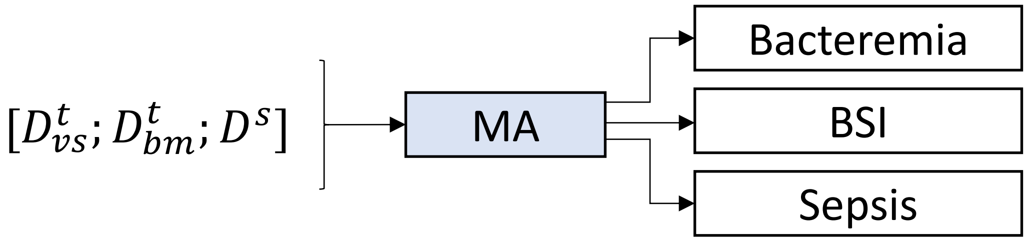


Results

Bacteremia	BSI	Sepsis
Running…	Running…	Running…

Approaches

Approach 2



Results

Bacteremia	BSI	Sepsis
Running…	Running…	Running…

Future work

- Implement multitask learning as a third approach
- Predict outcomes at T time steps look forward
- Evaluate different window sizes and sliding intervals (look backward)
- Assess interpretable architectures to identify the most relevant variables for the prediction task

THANK YOU FOR YOUR ATTENTION!

For any further doubt or suggestion

Óscar Escudero-Arnanz, oscar.escudero@urjc.es
PhD Student
Department of Signal Theory and Communications, Telematics and Computing Systems
Rey Juan Carlos University

URJC

Features considered

Biochemical markers

1. Albumin (alb) [g/dL]
2. Alkaline phosphatase (alp) [IU/L]
3. Alanine aminotransferase (alt) [IU/L]
4. Aspartate aminotransferase (ast) [IU/L]
5. Band form neutrophils (bnd) [%]
6. Base excess (be) [mEq/L]
7. Basophils (basos) [%]
8. Bicarbonate (bicar) [mEq/L]
9. Bilirubin direct (bildir) [mg/dL]
10. Blood urea nitrogen (bun) [mg/dL]
11. Calcium (ca) [mg/dL]
12. Carboxyhemoglobin (hbco) [-]
13. Chloride (cl) [mEq/L]
14. CO_2 partial pressure (pcO2) [mmHg]
15. Creatine kinase (ck) [IU/L]
16. Creatine kinase MB (ckmb) [ng/mL]
17. Creatinine (crea) [mg/dL]
18. C-reactive protein (crp) [mg/L]
19. Endtidal CO_2 (etcO2) [mmHg]
20. Eosinophils (eos) [%]

20. Eosinophils (eos) [%]

21. Erythrocyte distribution width (rdw) [%]
22. Erythrocyte sedimentation rate (esr) [mm/hr]
23. Fibrinogen (fgn) [mg/dL]
24. Glucose (glu) [mg/dL]
25. Hematocrit (hct) [%]
26. Hemoglobin (hgb) [g/dL]
27. Lactate (lact) [mmol/L]
28. Lymphocytes (lymph) [%]
29. Magnesium (mg) [mg/dL]
30. Mean cell hemoglobin (mch) [pg]
31. Mean corpuscular hemoglobin concentration (mchc) [%]
32. Mean corpuscular volume (mcv) [fL]
33. Methemoglobin (methb) [%]
34. Neutrophils (neut) [%]
35. O_2 partial pressure (po2) [mmHg]
36. Partial thromboplastin time (ptt) [sec]
37. Phosphate (phos) [mg/dL]
38. pH of blood (ph) [-]
39. Platelet count (plt) [K/ul]
40. Potassium (k) [mEq/L]

40. Potassium (k) [mEq/L]

41. Prothrombin time (inrpt) [-]
42. Prothrombine time (pt) [K/ul]
43. Red blood cell count (rbc) [m/uL]
44. Sodium (na) [mEq/L]
45. Total bilirubin (bili) [mg/dL]
46. Total CO_2 (tco2) [mEq/L]
47. Troponin I (tri) [ng/mL]
48. Troponin T (tnt) [ng/mL]
49. White blood cell count (wbc) [K/ul]
50. Respiratory rate (resp) [insp/min]
51. Endtidal CO_2 (etco2) [mmHg]
52. O_2 partial pressure (po2) [mmHg]
53. CO_2 partial pressure (pco2) [mmHg]
54. Urine output (urine) [mL]
55. Duration of ventilation (vent_dur) [sec]

Features considered

Vital signs

1. mean arterial pressure (map) [mmHg]
2. heart rate (hr) [bpm]
3. temperature (temp) [C]
4. systolic blood pressure (sbp) [mmHg]
5. diastolic blood pressure (dbp) [mmHg]
6. oxygen saturation (o2sat) [%]
7. calcium ionized (cai) [mmol/L]

Statics

1. age [years]
2. sex [-]
3. weight [kg]
4. height [cm]

Features considered

Biochemical markers					Vital signs	Statics
Albumin (**alb**) [g/dL]	alkaline phosphatase (**alp**) [IU/L]	alanine aminotransferase (**alt**) [IU/L]	aspartate aminotransferase (**ast**) [IU/L]	band form neutrophils (**bnd**) [%]	mean arterial pressure (**map**) [mmHg]	age [years]
base excess (**be**) [mEq/L]	Basophils (**basos**) [%]	Bicarbonate (**bicar**) [mEq/L]	bilirubin direct (**bildir**) [mg/dL]	blood urea nitrogen (**bun**) [mg/dL]	heart rate (**hr**) [bpm]	sex []
Calcium (**ca**) [mg/dL]	Carboxyhemoglobin (**hbco**) [-]	Chloride (**cl**) [mEq/L]	CO2 partial pressure (**pcO2**) [mmHg]	creatine kinase (**ck**) [IU/L]	temperature (**temp**) [C]	weight [kg]
creatine kinase MB (**ckmb**) [ng/mL]	Creatinine (**crea**) [mg/dL]	C-reactive protein (**crp**) [mg/L]	endtidal CO2 (**etcO2**) [mmHg]	Eosinophils (**eos**) [%]	systolic blood pressure (**sbp**) [mmHg]	height [cm]
erythrocyte distribution width (**rdw**) [%]	erythrocyte sedimentation rate (**esr**) [mm/hr]	Fibrinogen (**fgn**) [mg/dL]	Glucose (**glu**) [mg/dL]	Hematocrit (**hct**) [%]	diastolic blood pressure (**dbp**) [mmHg]	
Hemoglobin (**hgb**) [g/dL]	Lactate (**lact**) [mmol/L]	Lymphocytes (**lymph**) [%]	Magnesium (**mg**) [mg/dL]	mean cell hemoglobin (**mch**) [pg]	oxygen saturation (**o2sat**) [%]	
mean corpuscular hemoglobin concentration (**mchc**) [%]	mean corpuscular volume (**mcv**) [fL]	Methemoglobin (**methb**) [%]	Neutrophils (**neut**) [%]	O2 partial pressure (**po2**) [mmHg]	calcium ionized (**cai**) []	
partial thromboplastin time (**ptt**) [sec]	Phosphate (**phos**) [mg/dL]	pH of blood (**ph**) [-]	platelet count (**plt**) [K/uL]	Potassium (**k**) [mEq/L]		
prothrombin time (**inrpt**) [-]	prothrombine time (**pt**) [K/ul]	red blood cell count (**rbc**) [m/uL]	respiratory rate (**resp**) [insp/min]	Sodium (**na**) [mEq/L]		
total bilirubin (**bili**) [mg/dL]	totcal CO2 (**tco2**) [mEq/L]	troponin I (**tri**) [ng/mL]	troponin t (**tnt**) [ng/mL]	white blood cell count (**wbc**) [K/uL]		