1.1–1.5

Problem definition

Understand the clinical background to accurately define the problem.

- Domain expertise (e.g. clinical pathway)
- Population (e.g., inclusion criteria)
- Gold standard (e.g., bacteremia, sepsis)
- Aim (e.g. diagnosis, prediction)

6.1–6.6

Miscellaneous

- Cost-sensitive analysis
- Data availability
- Code availability
- Model availability
- Live demo

5.1–5.7

Deployment

- Legal and regulatory compliance
- Resources required
- Communication and transparency
- Establish a feedback loop
- Deployment and integration (e.g. human centred)
- Continuous monitoring and updates
- Patient diversity and bias

2.1–2.4

Data collection

- Relevance
- Availability
- Quality
- Frequency of updates
- Resource requirements
- Setting
- Invasivness
- Turnaround time
- Cost

2.1–2.4

Data sources

- History (e.g. comorbidities)
- Demographics (e.g., age, gender)
- Management (e.g., ward, LOS)
- Organ support (e.g., ventilation)
- Symptoms (e.g., rash, bleeding)
- Vital signs (e.g., BT, HR, RR, SBP)
- Biomarkers (e.g., WBC, PLT, CRP)
- Medication (e.g., antimicrobials)
- Other (e.g., blood culture, X-Ray)

2.5

Data leakage

- Target: information of the targed leaked into the training process
- Time: future observations used to predict past events

3.2 Start with a **statistical** approach using descriptive statistics, correlation analysis, clinical scoring sytems and/or survival analysis.

3.3 Continue with **traditional** machine learning approaches which are well established, facilitating interpretability and adoption (e.g. LR, DTC, RFC, SVM)

3.4 Explore **time series** approaches when data volume and quality allows and define clearly look-back and look-ahead settings (e.g. ARIMA, GRU, RNN, LSTM, ATTN)

3.5 Will an **ensemble** boost performance?

3.6–3.10

Consider **aspects** such as…

- Complexity
- Interpretability
- Adoption
- Adaptation
- Data requirements (volume, quality)
- Computational resources
- Tendency to overfit
- Consider mplementation
- Data governance and privace
- Data collection and storage
- Data anotation, labeling, ...

4.11–4.16

Performance evaluation

- Standard metrics (e.g., ROC, PR, SENS, SPEC)
- Further analysis
  - cut-off thresholds
  - feature importance
  - behaviour for different subpopulations (e.g., age, ward, ethnicity)
  - behaviour for different prevalence levels
  - behaviour for different levels of missing data
- Domain specific metrics and clinical relevance (e.g., alert fatigue)
- Understand the economic impact (e.g., costs)
- Fairness, bias and ethical considerations

4.1–4.2

Data cleaning

- Consider the study design requirements
- Discard duplicates, errors, ...
- Discard using the domain knowledge
- Discard outliers (e.g., physiological range)

4.3

Class imbalance

Acknowledge, describe and/or address class imbalance. Use appropriate metrics to evaluate model performance.

Data scaling

Consider the characteristics of your data and the algorithm's requirements before deciding whether and how to scale the features.

Imputation

4.4

- Cross-sectional data (e.g., median)
- Timeseries (e.g., ffill + median)
- Irrugularly sampled timeseries (MGP)

4.8–4.10

Model configuration

- Regularization
- Normalization
- Hyperparameter tunning: employ automated methods such as grid-search or bayesian optimization.

4.7

Data splitting

- Split the data into train, validation and a hold-out test sets.
- Consider time-aware splits to replicatemodel's usage as it would occur in a clinical setting.
- Use cross-validation

Feature engineering

4.6

- Domain knowledge (e.g., X/Y ratio)
- Categorization (e.g., low, medium, high)
- Stat. measures (e.g., min, max, range)
- Stat. momentums (e.g., mean, std)
- Temporal differences (e.g., t, t1–t, t2–t)
- Rolling statistics (e.g., mean HR over 1h)
- Time since the last event (e.g.,)
- Frequency encodings (e.g., DFT)
- Complex abstractions
- Dimensionality reduction (e.g., PCA)
- Indicator variables (e.g., missing mask)