Predicting the Clinical Outcomes of MRSA Infection

Presenter: Bella Yang (1st year PhD student)
Supervisors: Prof. Rita Ng and Prof. Margaret Ip

Joint postgraduate seminar

24th November 2023

Outline

- **General introduction**
  - The context
  - The motivation
  - The pipeline of model development
- **Examples**
  - Statistical models
  - Machine learning models
- **Current challenges**
  - Model development
  - Application

Staphylococcus aureus

- ***Staphylococcus aureus*** is a common bacteria found on skin and in nasal.

- **Infections**: mild to severe
  - Skin and soft tissue infection.
  - Respiratory tract infection.
  - Renal/urinary tract infection.
  - Abdominal infection.
  - Catheter-related infection.
  - Bacteraemia.
  - Bloodstream infection (sepsis).
  - Central nervous system infection.

(Hanberger et al, 2011)

MRSA

- ***MecA* gene (transmit through mobile genetic elements)**
  - Modify or overexpress penicillin-binding proteins (PBPs; peptidoglycan transpeptidase on cell wall) - PBP-2a

    👉**Reduce avidity to most of the β-lactams** (e.g. oxacillin or cefoxitin).

- **Methicillin-resistant *Staphylococcus aureus* (MRSA)**
  - Healthcare-associated MRSA (HA-MRSA)
  - Community-associated MRSA (CA-MRSA)
  - Livestock-associated MRSA (LA-MRSA)

(Foster and Geoghegan, 2015, Lee et al., 2018)

The burden of MRSA infection

- The second leading pathogens for deaths associated with resistance.
- In 2019, MRSA caused more than 100,000 deaths attributable to AMR globally.
- The prevalence of MRSA resistance varied across different regions and countries.
  - Highest: north Africa and the middle east countries (> 60%).
  - Lowest: several Europe and sub-Saharan countries (< 5%).

(Murray et al, 2022)

The motivation of developing prediction models for the clinical outcomes

- Early identification of high-risk patients.
  - Early warning systems.
  - Develop or improve risk assessment tools .
- Preventing complications and reducing mortality.
- Optimizing antibiotic prescription.
- Efficient allocation of healthcare resources.

Classical statical modelling: pipeline

Dataset:
- **Collection**
- **Exploration**
- **Cleaning**
- **Preparation**

- **Assumptions check**
- **Fitting an existing statistical model**

E.g. predictor (variable) selection for logistic regression model:
- Stepwise
- Backward stepwise
- Forward stepwise
- All-subsets

Validity check:
- Goodness-of-fit
- Model diagnosis

- ***Evaluation with external data (observed *VS* predicted value)**
  - e.g. public databank

- Widely used in biological research.
- Choose a specific model and adjust. parameters based on our current knowledge.
- Advantage: High interpretability.

(Bzdok et al, 2018)

Machine learning modelling: pipeline

Dataset:

- **Collection**
- **Exploration**
- **Cleaning**
- **Preparation**
- **Splitting ('train-test')**

Model training:

- **Feature selection**
- **Create new features**
- **ML algorithm selection**

Feature selection strategies (interpretability and accuracy):

- Domain-specific knowledge.
- Data-driven: using statistics and/or machine learning techniques.

Model performance evaluation:

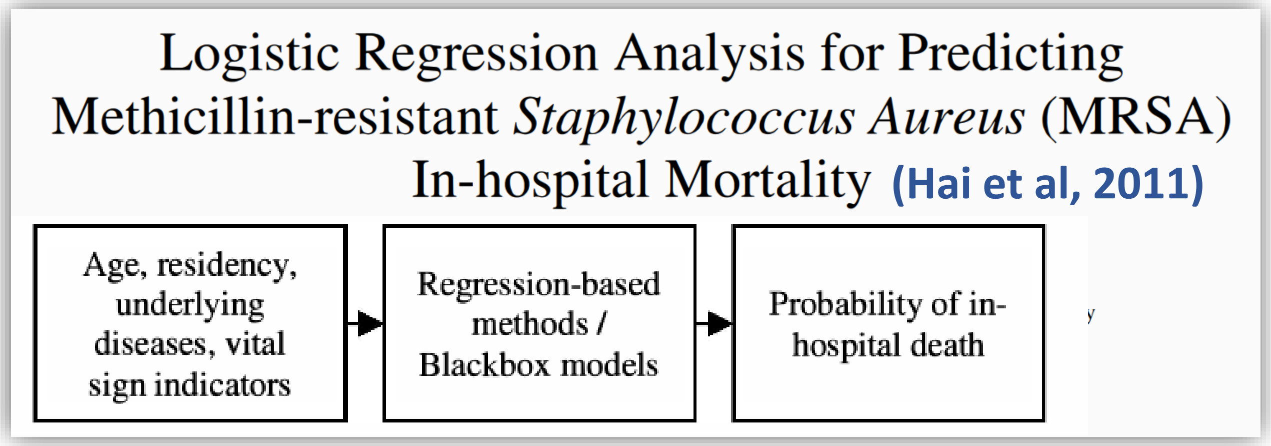
- Accuracy
- Precision
- Recall
- F1-score
- AUC
- **External data validation (i.e. test/ "outgoing Quality Control").**
- **Comparison between models.**
- **Clinical traits.**

Application:

- **Evaluation of the adoption and integration into clinical workflows by end-users** (i.e. physician and nurses).

(Luz et al 2020; Neumann et al, 2017; Figueroa Barraza et al, 2021)

Example 1: Statistical models



MRSA infection in Queen Mary Hospital

- **Objective:** predict the in-hospital mortality.
- **Data source:** clinical management system (2006 – 2010)
- **Method** (1,762 patients)
  - Logistic regression model.
    - Variables were selected based on Chi-square test and Welch two sample t-test ($p<0.1$).
- **Results**:
  - z = -3.49 + 0.01*age - 0.71* Residency + 0.52*Solid tumor +1.03*Hemic malignancy + 0.76*COAD + 0.94*Dementia + 0.52*PLT + 0.55*Lymphocyte + 0.53*Urea + 0.48*ALP
  - (Probability of death) $f(z) = 1/(I + e^{-Z})$

Future study suggested by the authors:

- “Possible direction is to make use of other data mining "blackbox" methods, such as k-NN (K-Nearest Neighbours) and SVM (Support vector Machine). These models also need further validation on their performance and feature selection”.

Example 2: Statistical models

MRSA blood stream infection in Hong Kong (1,133 patients)

- **Objective:** describe the characteristics of 30-day mortality rate.
- **Data source:** electronic medical records - 26 Hong Kong public hospitals
- **Method:**
  - Logistic regression model.
    - Backward stepwise elimination.
    - The potential associations (*P* value ≤.1).

- **Results**: predictors of mortality:

Final Model From Multivariate Analysis

Variable	Odds Ratio (95% CI)	*P* Value
Older age (>79 years)	1.436 (1.099-1.877)	.008
Underlying chronic lung disease	1.671 (1.101-2.536)	.016
Skin and soft-tissue infection with MRSA	0.474 (0.296-0.759)	.002
Prior hospitalization	2.019 (1.244-3.279)	<.001
Long-term dialysis	0.415 (0.263-0.654)	<.001

- Odd ratio < 1: associated with lower risk.
- No collinearity was identified in predictors, and there was no significant interaction term found.

Disease Burden, Characteristics, and Outcomes of Methicillin-Resistant *Staphylococcus aureus* Bloodstream Infection in Hong Kong

201
Rep
sagepub.com/jo
DOI: 10.117
journals.sa

(You et al, 2017)

Joyce H. S. You, DPharm, BCPS-AQ ID[1], Kin-wing Choi, MBChB, FRCP (Edin)[2], Tin-yau Wong, MBBS, MPH[2], Margaret Ip, BM, MSc[3], Wai-kit Ming, MD, MPH[1], Rity Yee-kwan Wong, BSN, MN[4], Sze-ngai Chan, MBBS[1], Hoi-tung Tse, MBBS[1], Carrie T. S. Chau, BSc, MPhil[1], and Nelson L. S. Lee, MBBS(HK), MD(CUHK)[4]

Example 3: Statistical models

Key predictors and burden of meticillin-resistant *Staphylococcus aureus* infection in comparison with meticillin-susceptible *S. aureus* infection in an Australian hospital setting (Miyakis et al, 2022)

S. Miyakis[a,b,*], S. Brentnall[c], M. Masso[d], G. Reynolds[b,e], M.K. Byrne[f],

Benefits of the study:
- *Antimicrobial stewardship*
- *Infection control practices*
- *Public awareness*

MRSA infection in Australia

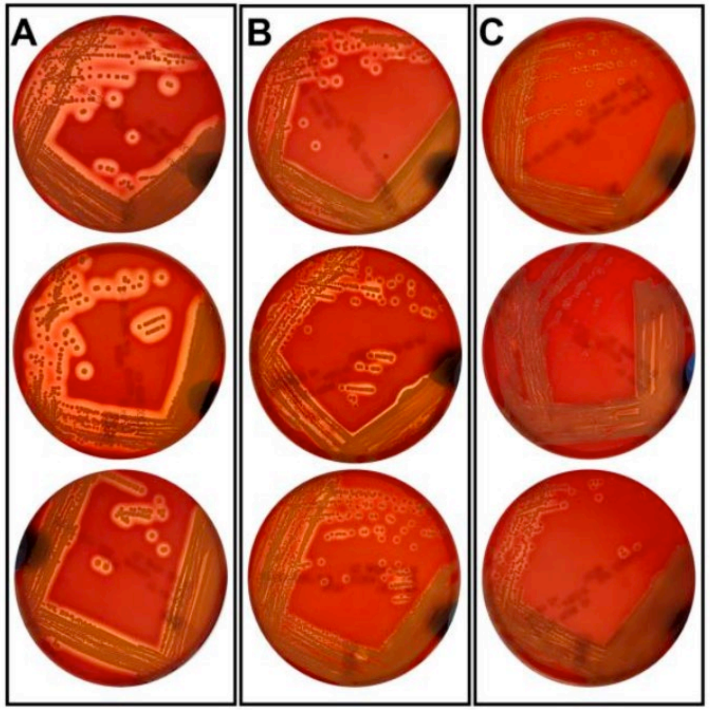
- **Objective:** to compare patients with MRSA and MSSA mortality and determine significant predictors of inpatient mortality.
- **Data source:** a non-identifiable databank established by the Centre for Health Research Illawarra Shoalhaven Population.
- **Method**: Cox proportional hazards model (5,897 patients) – the hazard ratio changed overtime.

- **Results**: predictors of survival probability of *S. aureus* infection (including MRSA and MSSA) - the first 375 days after admission ($P < 0.05$):
  - MRSA (compared to MSSA)
  - Older age
  - Male sex
  - Higher comorbidity score
  - Admission to a surgical ward was associated with lower inpatient mortality

Example 4: Machine learning models

Machine Learning with Alpha Toxin Phenotype to Predict Clinical Outcome in Patients with *Staphylococcus aureus* Bloodstream Infection (Beadell et al, 2023)

Brent Beadell [1,†], Surya Nehra [2,†], Elizabeth Gusenov [1], Holly Huse [3] and Annie Wong-Beringer [1,4,*]



MRSA bloodstream infection

- **Motivation:** Enable precision infectious disease therapeutics.
  - Alpha toxin-mediated thrombocytopenia (host-immune response).
    - Bacteria virulence factor production (i.e. alpha toxin (Hla))
- **Objective:** Predict thrombocytopenia on day 4 (platelet count < 150 * 10^9 /L) and 30-day mortality.
- **Data source:** Patients' medical records and REDCap electronic data capture tools.
- **Method:**

- **539 smartphone blood agar images.**
- **Day 1 platelet count of 229 patients.**

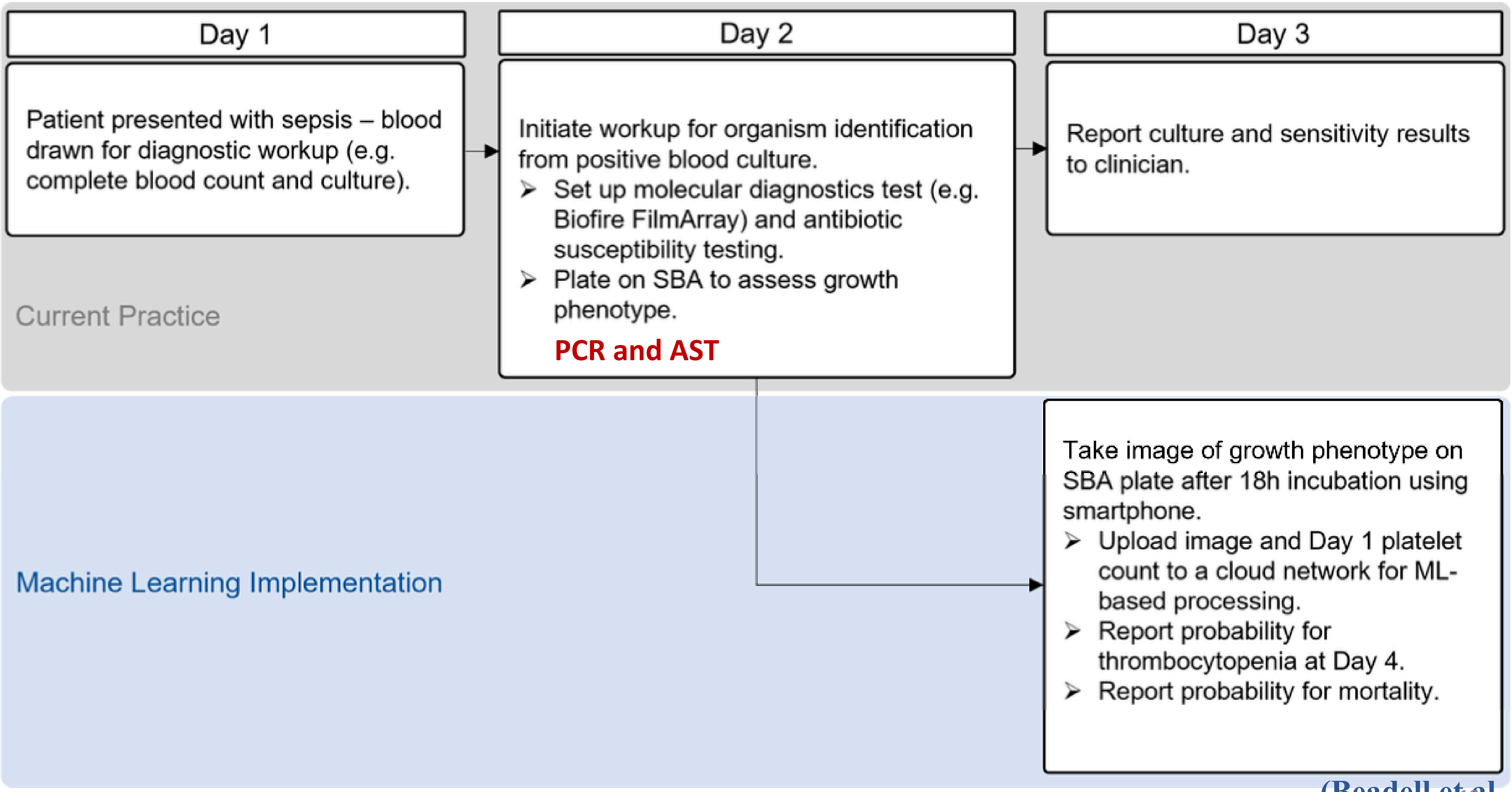
→ **Convolutional Neural Network** →

Outcomes:
- **30-days mortality.**
- **Day 4 thrombocytopenia.**

→

5-fold cross-validation:
- Day 4 platelet count prediction: AUC = 0.920 (strong).
- 30-day mortality: AUC = 0.711 (modest).

Integrate into the clinical workflow:



(Beadell et al, 2023)

Example 5: Machine learning models

Machine learning models for predicting in-hospital mortality in patient with sepsis: Analysis of vital sign dynamics

(Cheng et al, 2022)

Chi-Yung Cheng[1,2], Chia-Te Kung[2], Fu-Cheng Chen[2], I-Min Chiu[1,2], Chun-Hung Richard Lin[1], Chun-Chieh Chu[2], Chien Feng Kung[3]* and Chih-Min Su[2]*

Motivation: build the early warning system model (ESM).

Objective: to predict the in-hospital death within 6 to 48 hours of admission.

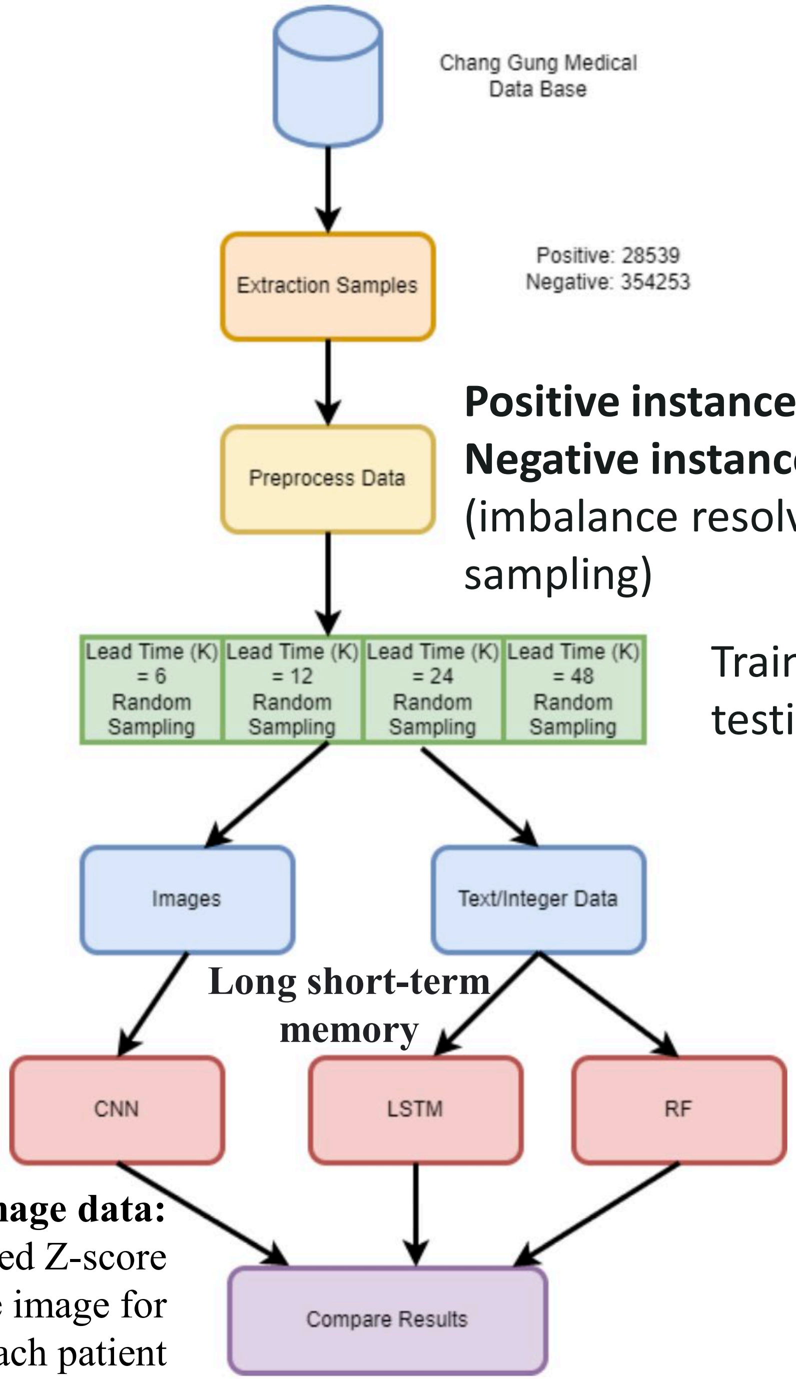
Data source: electronic database records of Chang Gung Medical Center.

Method:

Features:

- Five vital signs: heart rate, respiratory rate, systolic blood pressure, diastolic blood pressure, and body temperature.
- Age and sex.

Validation: 5-fold cross-validation and an extra validation with reserved data.



Positive instances: 19,434; **Negative instances:** 194,646 (imbalance resolved by random sampling)

Training : validation : testing sets = 6:2:2

Long short-term memory

Random forest

Image data: normalized Z-score grayscale image for each patient

Current challenges

- More attention should be given to the calibration (i.e. the reliability of risk predictions).
  - The current focus is primarily on the discrimination performance with traditional index (e.g. accuracy and precision)
- Integration of AI into clinical settings:
  - Identify which algorithms have the best performance for different types of prediction problems.
  - Who will be responsible for AI (i.e. algorithm bias/ errors)?

(Christodouloua et al, 2019; Luz et al, 2020)

Thank you for your attention!