# Early Sepsis Risk Prediction in ICU

## Using Machine Learning Models

**By: Yodha Pranata, RN**

**FINAL PROJECT**
as Data Scientist

# Introduction

## Work Experiences

- Critical Care Nurse
  Public Hospital (2019 - 2024)
- Career Center Coordinator
  University of Kepanjen (2018 - 2019)

## Education

**Bachelor of Data Science**
Universitas Terbuka (2024 - Present)

**Master of Critical Care**
Universitas Padjadjaran (2024 - Present)

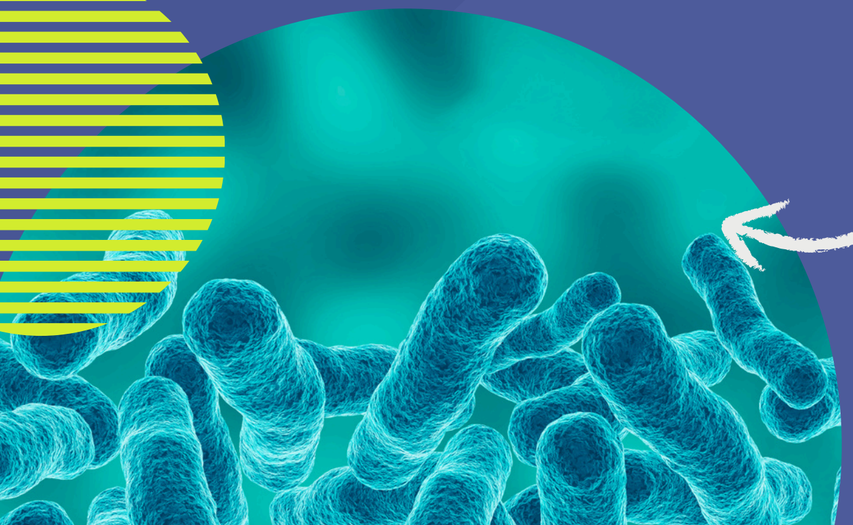
**Bachelor of Nursing Science & Nursing Profession**
Universitas Brawijaya (2012 - 2017)


# Yodha Pranata

**LinkedIn:**
linkedin.com/in/yodha-pranata

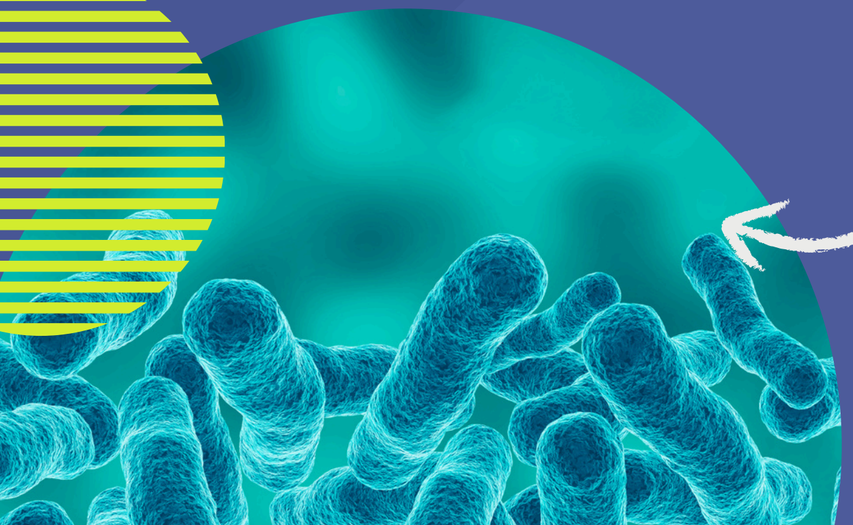
# Project Background

- **Sepsis** is a life-threatening organ dysfunction caused by a dysregulated response to infection.
- **Global burden is huge**: a 2020 analysis estimated **48.9 million cases** and **11 million deaths** worldwide in 1990-2017 (about **1 in 5 deaths worldwide** are linked to sepsis).
- Mortality in severe sepsis remains high (**often >20% in hospitals**), and outcomes worsen in low-resource settings.
- **Time is critical:** each hour of delayed recognition and treatment increases organ failure and death risk, so sepsis is treated as a medical emergency.



# Project Background

- **Early sepsis** is **hard to detect** because symptoms are non-specific and there is no single definitive early test.
- Standard tools such as **SIRS** helps early screening but is **nonspecific**, while **SOFA/qSOFA** mainly **capture** established **organ dysfunction** and prognosis, so early sign sepsis can be missed.
- **ICU data are high-volume** and **complex** (vitals, labs, EMR), making **pattern detection difficult** by manual review alone.
- **Machine learning** can integrate multi-variable patterns from the first 24 hours of ICU data to stratify patients' early sepsis risk, **helping clinicians prioritize monitoring** and **initiate sepsis bundles** sooner, which is **associated with better outcomes.**

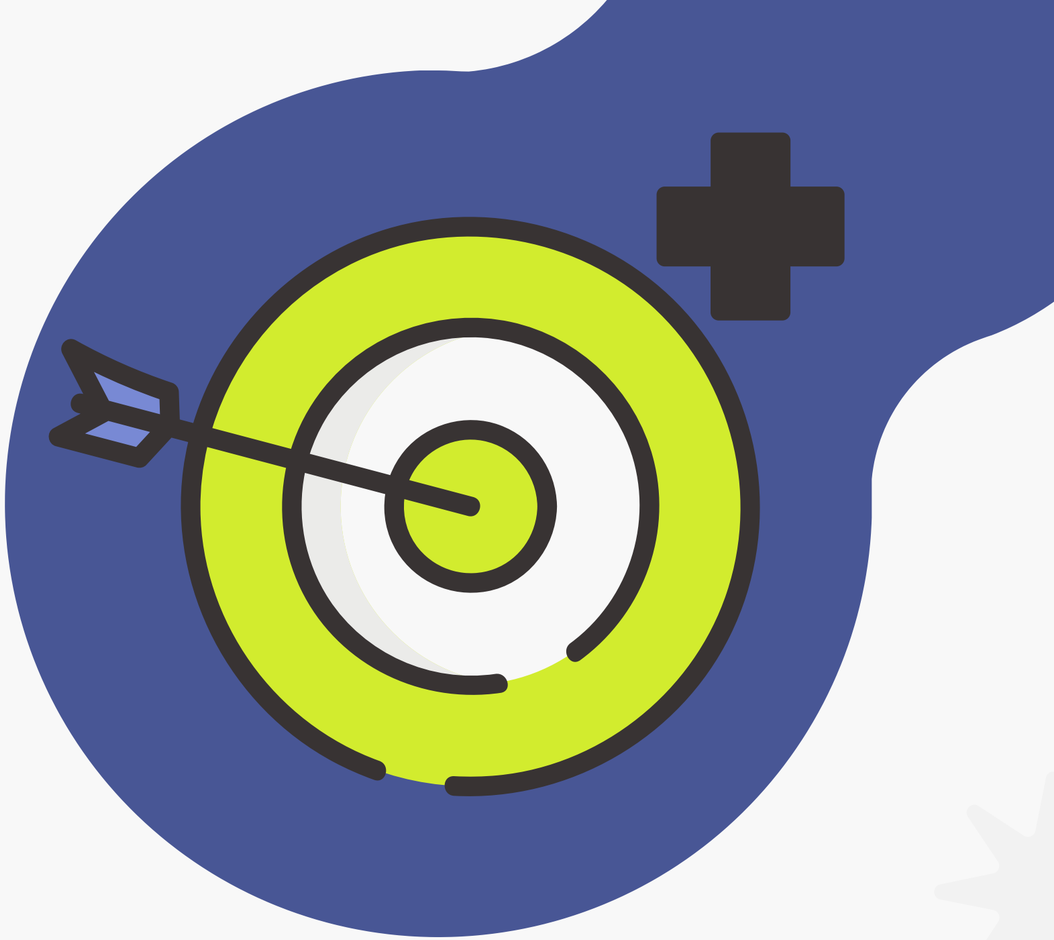


# Problem Statement

How accurately can **early sepsis risk** in ICU patients be **predicted** using routinely collected clinical **data from the first 24 hours** in ICU?

# Objective

This project aims to **develop** and **validate** a **predictive model** for early sepsis risk in ICU patients using routinely collected clinical data from the first 24 hours in ICU.



# Data Understanding

- **Data source:** Public dataset from **Kaggle**, originally used in the *PhysioNet/Computing in Cardiology Challenge 2019*.
- Hourly **time-series format**
- **Mixed feature types** (Vital Sign, laboratory results, and Basic Information Patients)

1,552,210
**rows**

42
**features**

40,321
**ICU patients**

| Row | Measurement | Description |
|---|---|---|
| 1 | HR | Heart rate (beats/min) |
| 2 | $O_2$Sat | Pulse oximetry (%) |
| 3 | Temp | Temperature (°C) |
| 4 | SBP | Systolic BP (mm Hg) |
| 5 | MAP | Mean arterial pressure (mm Hg) |
| 6 | DBP | Diastolic BP (mm Hg) |
| 7 | Resp | Respiration rate (breaths/min) |
| 8 | $Etco_2$ | End tidal carbon dioxide (mm Hg) |
| 9 | BaseExcess | Excess bicarbonate (mmol/L) |
| 10 | $Hco_3$ | Bicarbonate (mmol/L) |
| 11 | $Fio_2$ | Fraction of inspired oxygen (%) |
| 12 | pH | pH |
| 13 | $Paco_2$ | Partial pressure of carbon dioxide from arterial blood (mm Hg) |
| 14 | $Sao_2$ | Oxygen saturation from arterial blood (%) |
| 15 | AST | Aspartate transaminase (IU/L) |
| 16 | BUN | Blood urea nitrogen (mg/dL) |
| 17 | Alkalinephos | Alkaline phosphatase (IU/L) |
| 18 | Calcium | Calcium (mg/dL) |
| 19 | Chloride | Chloride (mmol/L) |
| 20 | Creatinine | Creatinine (mg/dL) |
| 21 | Bilirubin direct | Direct bilirubin (mg/dL) |
| 22 | Glucose | Serum glucose (mg/dL) |
| 23 | Lactate | Lactic acid (mg/dL) |
| 24 | Magnesium | Magnesium (mmol/dL) |
| 25 | Phosphate | Phosphate (mg/dL) |
| 26 | Potassium | Potassium (mmol/L) |
| 27 | Bilirubin total | Total bilirubin (mg/dL) |
| 28 | TroponinI | Troponin I (ng/mL) |
| 29 | Hct | Hematocrit (%) |
| 30 | Hgb | Hemoglobin (g/dL) |
| 31 | PTT | Partial thromboplastin time (s) |
| 32 | WBC | Leukocyte count (count/L) |
| 33 | Fibrinogen | Fibrinogen concentration (mg/dL) |
| 34 | Platelets | Platelet count (count/mL) |
| 35 | Age | Age (yr) |
| 36 | Gender | Female (0) or male (1) |
| 37 | Unit 1 | Administrative identifier for ICU unit (medical ICU); false (0) or true (1) |
| 38 | Unit 2 | Administrative identifier for ICU unit (surgical ICU); false (0) or true (1) |
| 39 | HospAdmTime | Time between hospital and ICU admission (hours since ICU admission) |
| 40 | ICULOS | ICU length of stay (hours since ICU admission) |
| 41 | SepsisLabel | For septic patients, SepsisLabel is 1 if $t \geq t_{sepsis} - 6$ and 0 if $t < t_{sepsis} - 6$. For nonseptic patients, SepsisLabel is 0. |

Clinical time series data provided for the Challenge: vital signs (rows 1–8), laboratory values (rows 9–34), demographics (rows 35–40), and outcome (row 41; only provided for the shared datasets).

# Data Preprocessing
## for EDA

### 1 Feature Aggregation

- **Raw data:** 1,552,210 hourly rows, 41 features
- **Split features** → Static | Vitals | Labs
- **Vitals missing** → per-patient *ffill/bfill*
- **Labs missing** → kept as-is (not measured hourly)
- **Aggregate first 24h**
  a. Vitals: mean / median / std / min / max
  b. Labs: mean / median / std / min / max / count
- **Output** → patient-level table **40,321 rows × 137 features**

### 2 Handling Duplicate

No Duplicate Data Found

### 3 Handling Missing Value

- **Vitals cols missing** → median imputation
- **Labs cols missing** → keep as NaN (likely not tested)

# Data Preprocessing
## for Machine Learning

### 1 Feature Aggregation

- **Raw data:** 1,552,210 hourly rows, 41 features
- **Split features** → Static | Vitals | Labs
- **Vitals missing** → per-patient *ffill/bfill*
- **Labs missing** → kept as-is (not measured hourly)
- **Aggregate first 24h**
  a. Vitals: mean / median / std / min / max
  b. Labs: mean / median / std / min / max / count
- **Output** → patient-level table **40,321 rows × 137 features**

SPLIT TRAIN & TEST

### 2 Handling Duplicate

No Duplicate Data Found

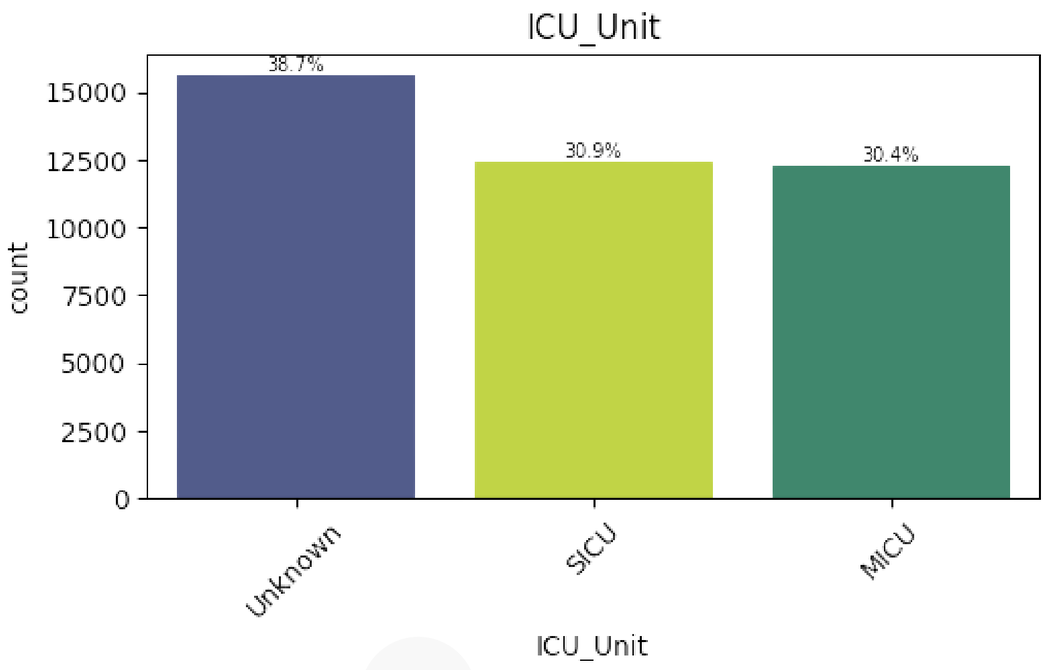
### 3 Handling Missing Value

- **Vitals cols missing** → median imputation
- **Labs cols missing** → keep as NaN (likely not tested)
- **Drop** → patient_id (identifier, non-predictive)

### 4 Outlier Screening

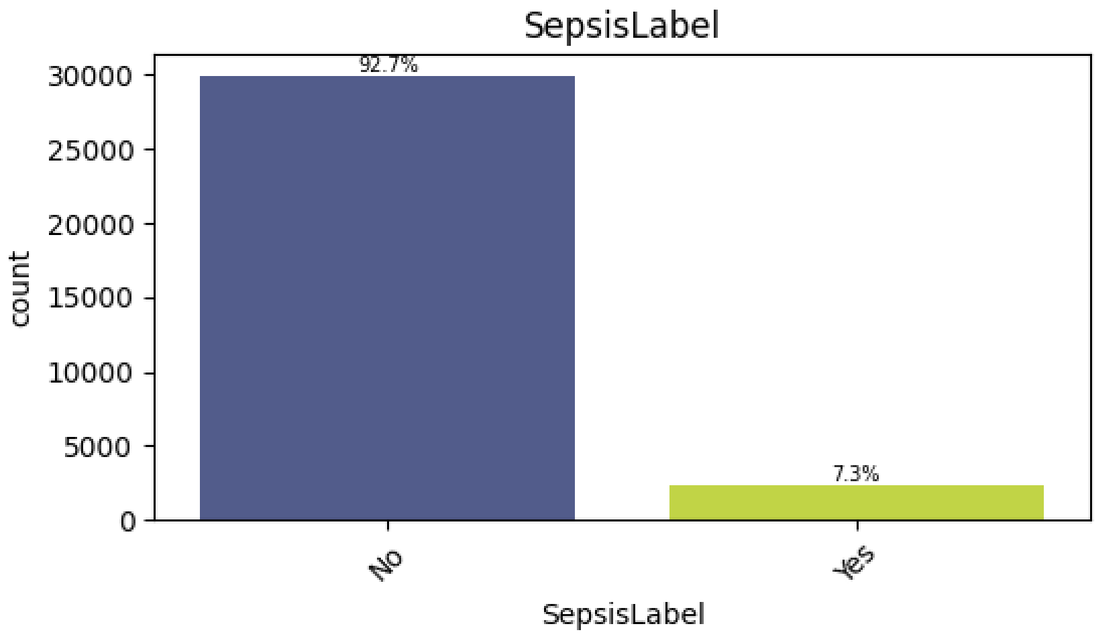
- **Outliers detected** → **retained**
- **Rationale** → values still within **physiological ranges** and **reflect true ICU critical conditions** relevant to sepsis prediction

# Exploratory Data Analysis



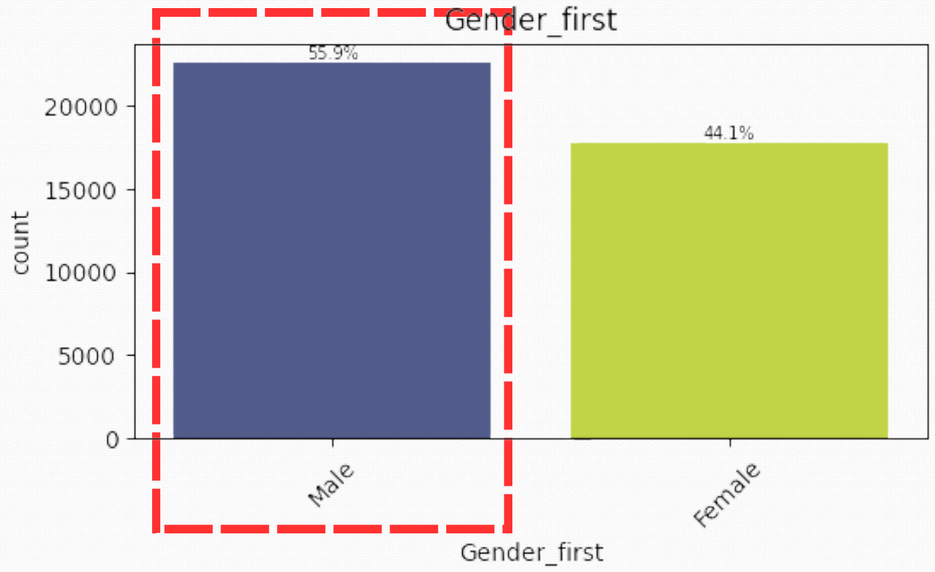

ICU unit labels are split into **Unknown** (38.7%), **SICU** (30.9%), and **MICU** (30.4%).

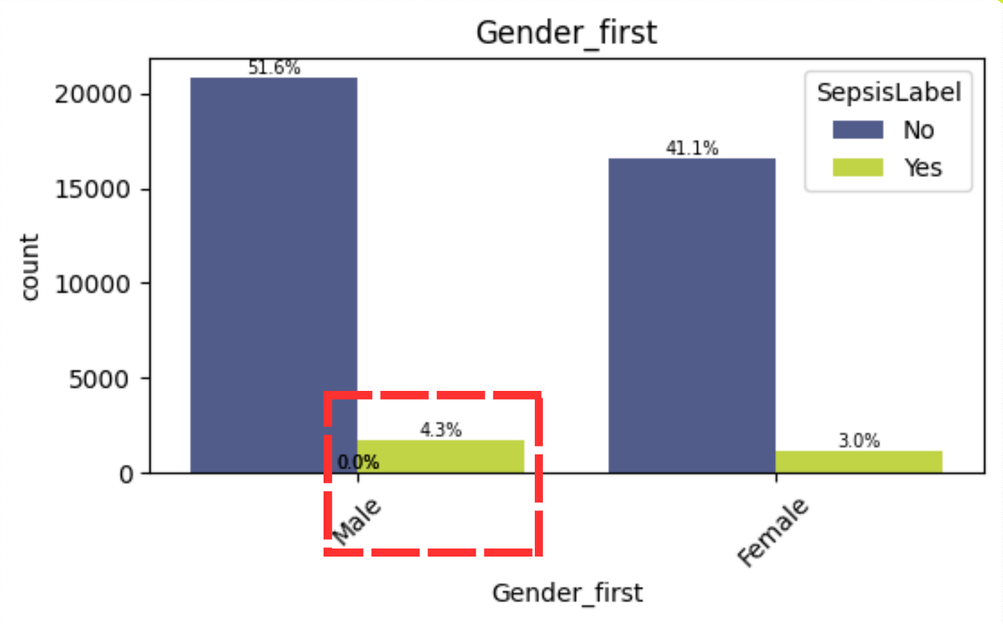
# Exploratory Data Analysis




- The target label is **highly imbalanced,** with **92.7% non-sepsis** and only **7.3% sepsis cases.**
- This skew indicates a strong class imbalance, so modeling should use strategies such as **class weighting** and report metrics beyond accuracy (e.g. AUPRC).

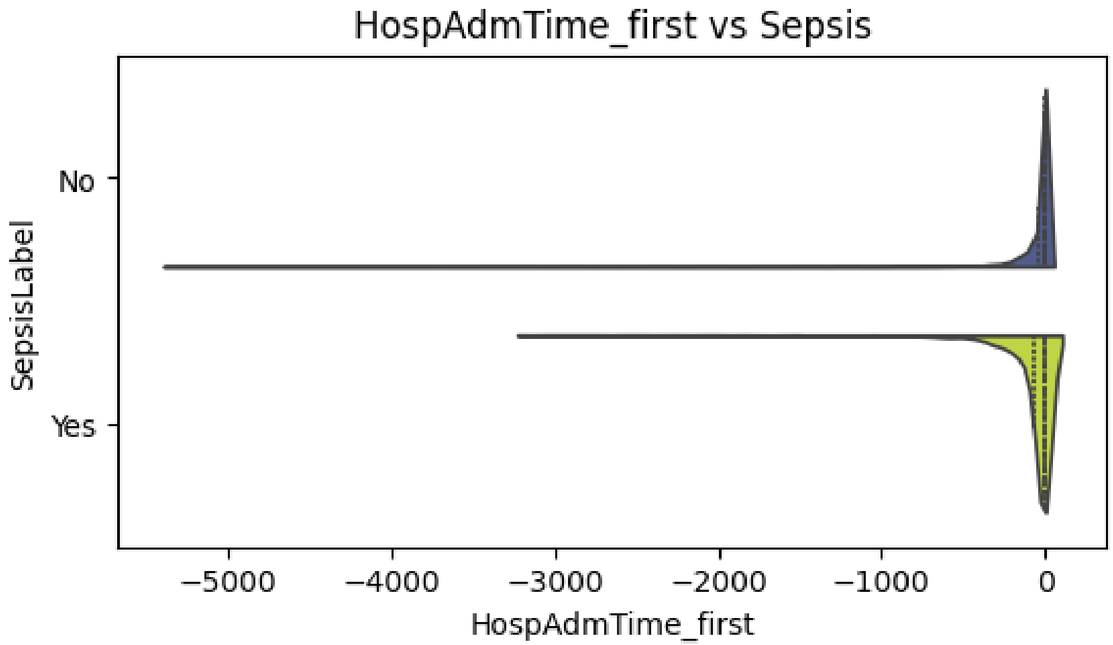
# Exploratory Data Analysis



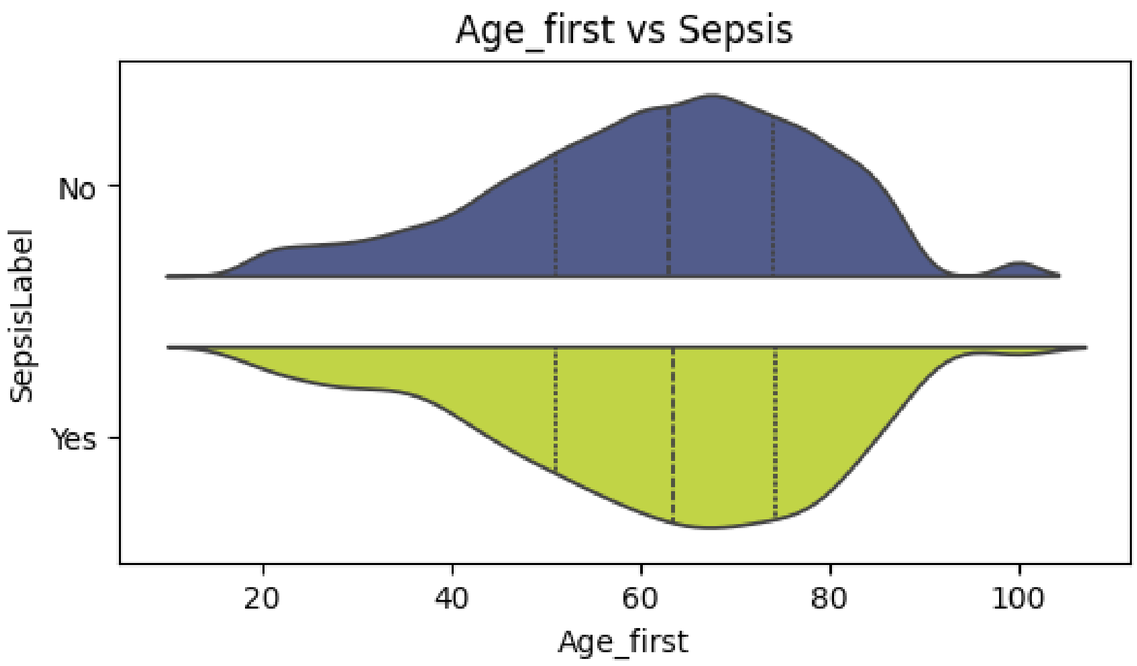

- **The dataset consists** of 55.9% males and 44.1% females.
- Sepsis remains rare but is slightly more common in **men: 4.3% of males** have sepsis versus **3% of females**. Non-sepsis accounts for 51.6% of the total sample in males and 41.1% in females.

# Exploratory Data Analysis




The plot shows that HospAdmTime_first is concentrated near 0 for both groups, meaning most **patients enter the ICU soon after hospital admission**. However, the **sepsis group** has a much **longer negative tail**, indicating **many sepsis cases had a substantially longer pre-ICU hospital stay**. This suggests prolonged hospitalization before ICU admission is associated with higher sepsis risk, likely **reflecting hospital-acquired infection** or delayed deterioration.

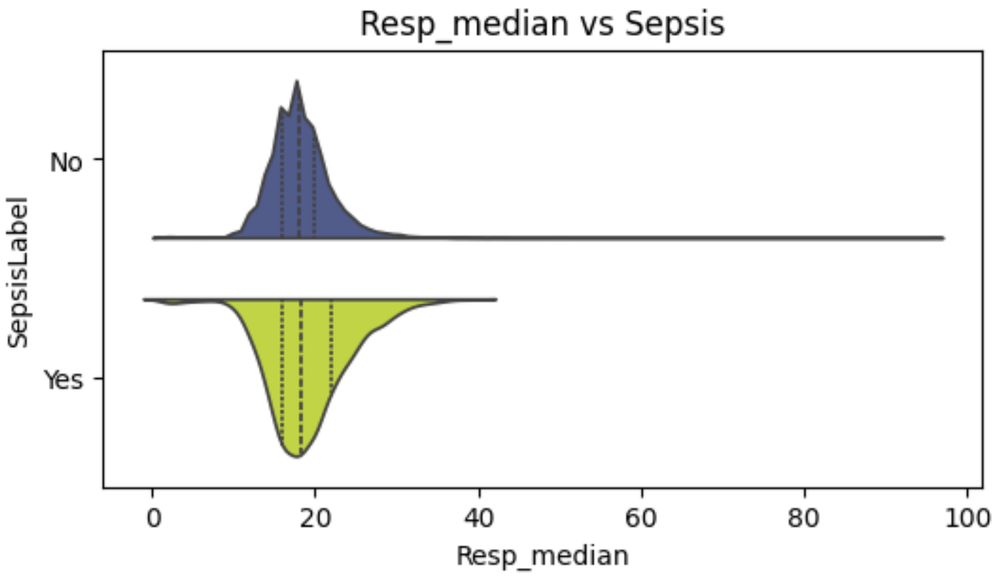
# Exploratory Data Analysis

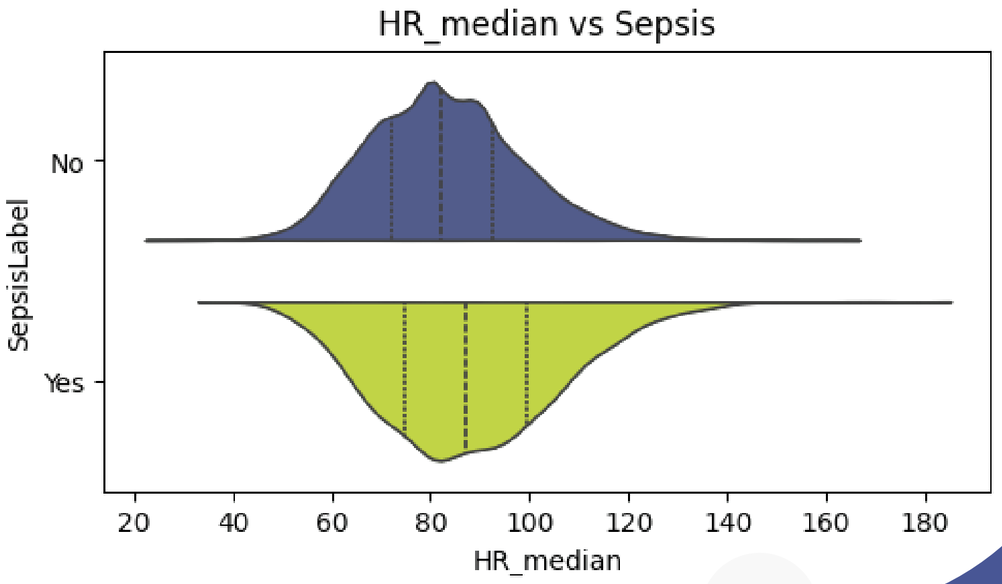


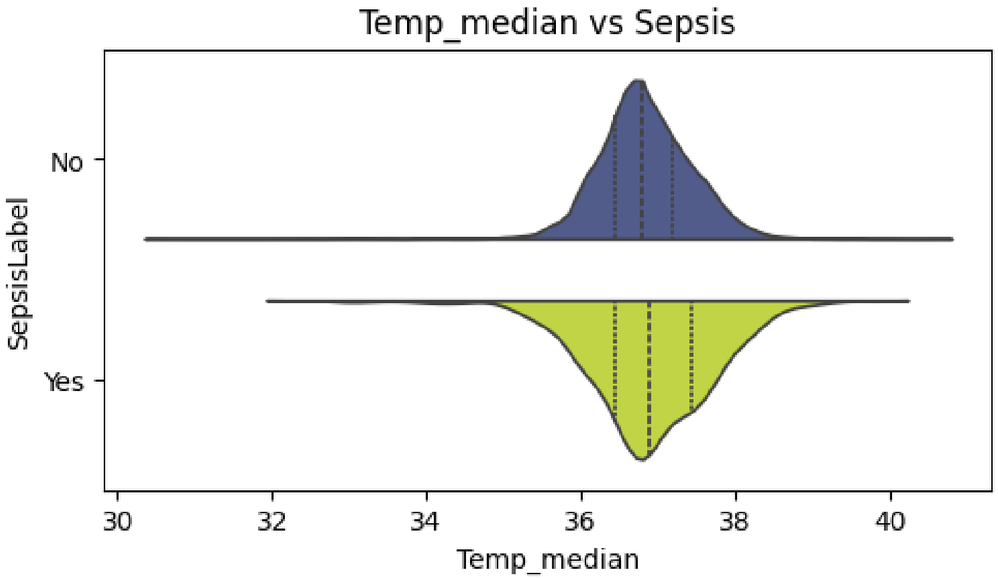

- **Most patients** cluster around **65–75 years**, indicating this is the dominant age band **in both groups**.
- The sepsis and non-sepsis age **distributions overlap** closely with no clear shift in median or spread.
- The **Mann–Whitney U test** $p = 0.12$ confirms that Age_first **does not differ significantly** between sepsis and non-sepsis patients in this dataset.

# Exploratory Data Analysis

# Vital Signs



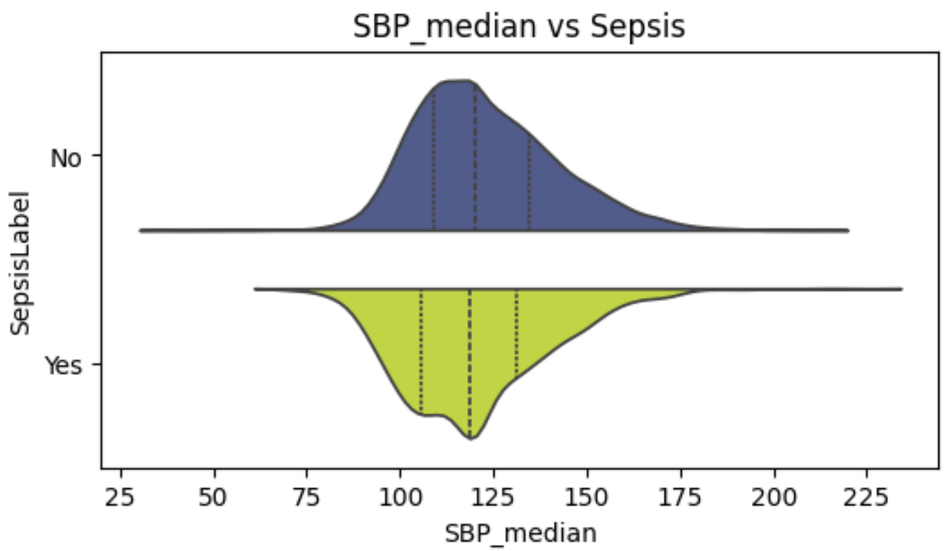



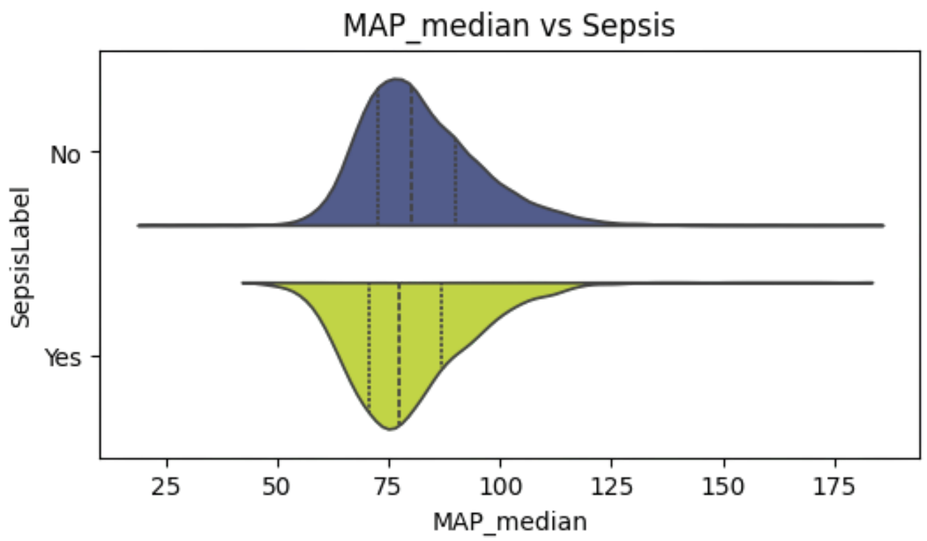



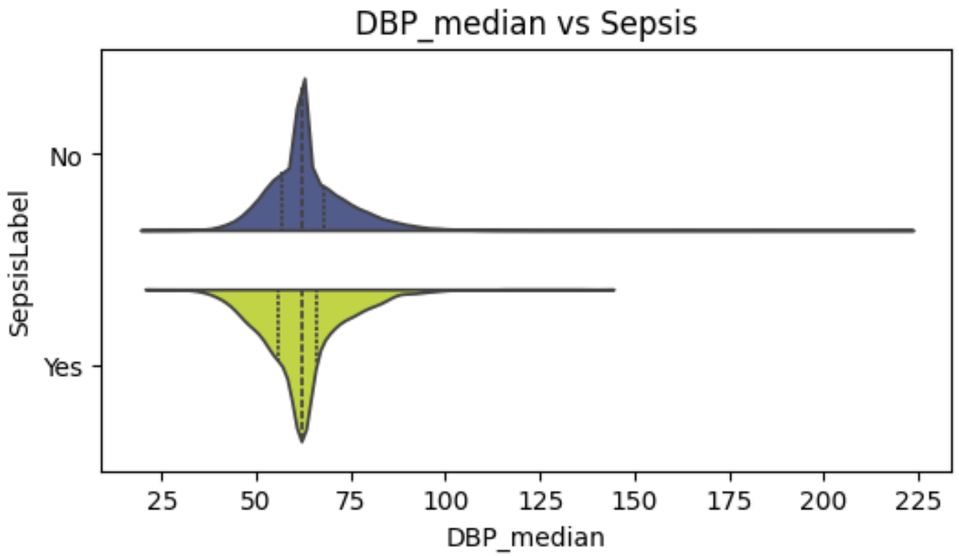

- The **sepsis group** shows **shifted** and **wider** distributions for all vitals, indicating early **physiologic instability**.
- **HR** and **RR** medians tend to be **higher in sepsis** patients, while temperature is slightly higher and more variable.
- **Mann–Whitney U** $p$ = .00 for HR, RR, and Temp confirms these differences are **statistically significant.**

# Exploratory Data Analysis

# Blood Pressures



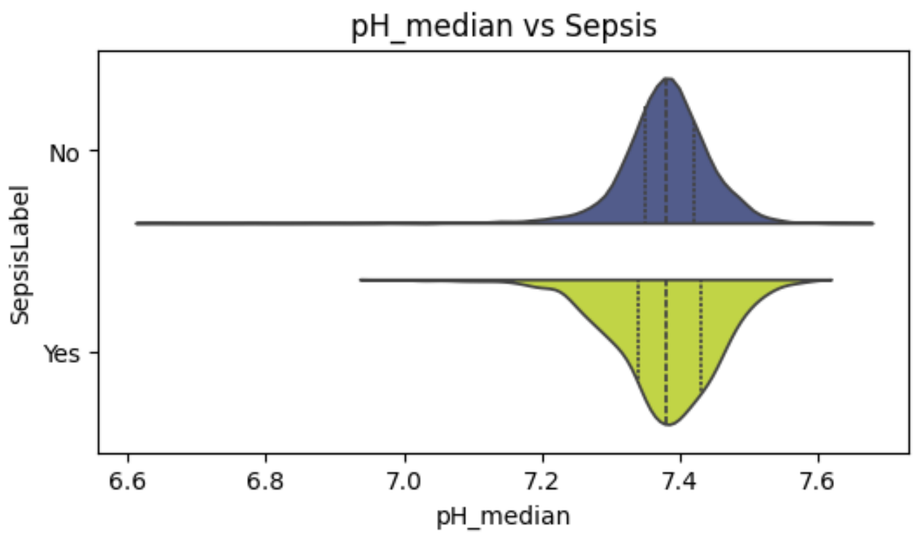



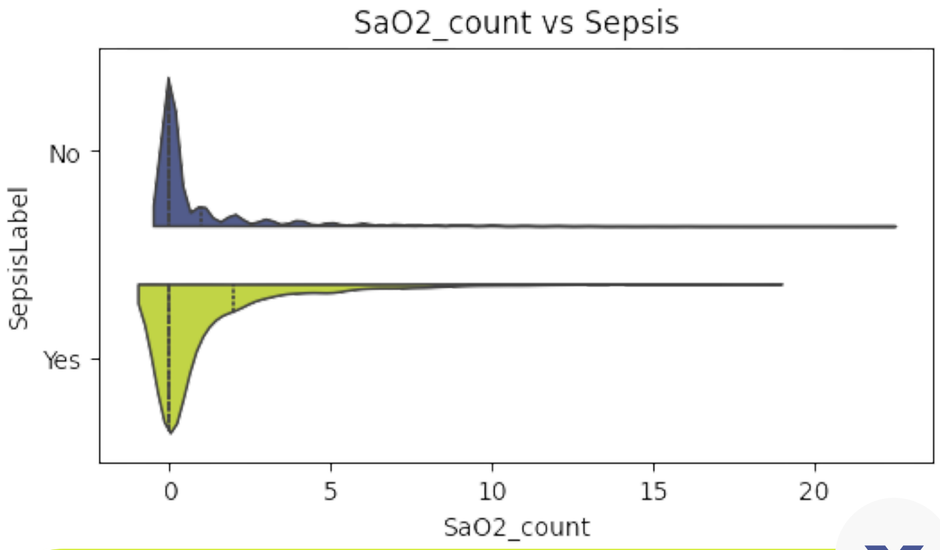



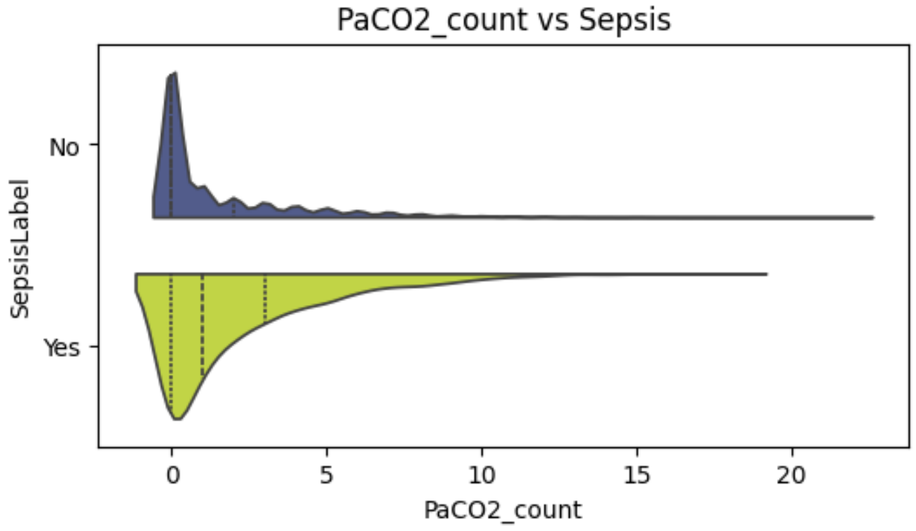

- The **sepsis group** has **lower median** MAP, SBP, and DBP, with curves shifted left.
- The **sepsis** distributions are wider, indicating early **hemodynamic instability/hypotension**.
- **Mann–Whitney U test** $p$ = .00 confirms these blood-pressure differences are **statistically significant.**

# Exploratory Data Analysis

# Blood Gas Analyses



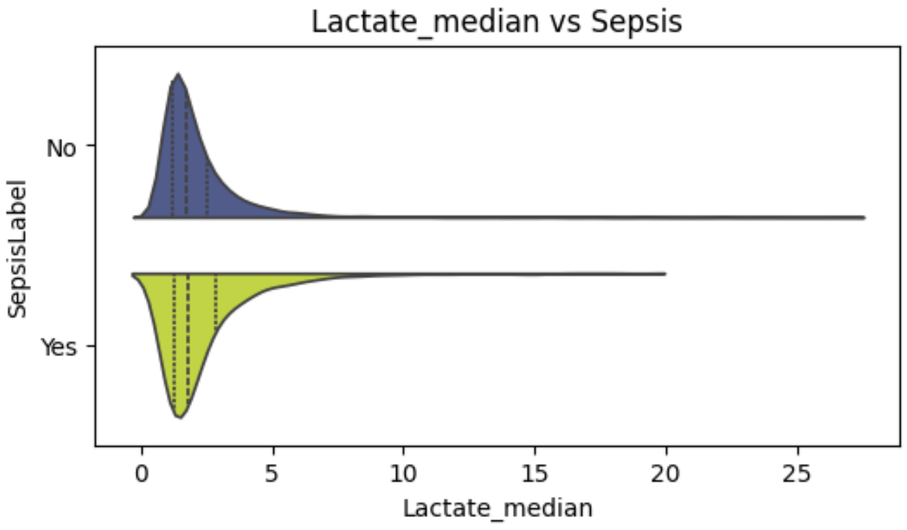



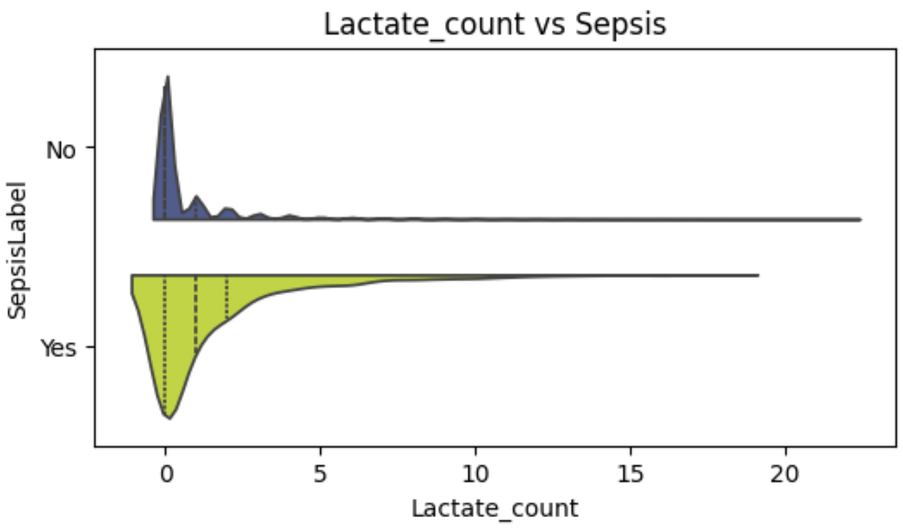




- **pH_median** is more variable and **slightly lower in the sepsis group**, indicating early **acid–base disturbance**.
- **PaCO2_count** and **SaO2_coun**t are **higher in sepsis,** suggesting **more frequent BGA checks** due to instability.
- **Mann–Whitney U** $p = 0.00$ confirms these differences are **significant.**

# Exploratory Data Analysis

# Lactate




- **Lactate_median** is **higher in sepsis**, indicating early hyperlactatemia linked to **hypoperfusion/shock**; non-sepsis stays lower and tighter.
- **Lactate_count** is **higher in sepsis**, reflecting **more frequent monitoring**.
- **Mann–Whitney U** $p = 0.00$ confirms both differences are **significant**.

# Clinically Retained Features

- These variables were **retained** a priori **based on critical-care expertise** and **sepsis pathophysiology**, and will next undergo **multicollinearity screening** (e.g., correlation) to remove redundant predictors while preserving clinical meaning.

**Retained 38 features :**

- **basic:** HospAdmTime_first
- **vital sign:** HR_median, HR_max, O2Sat_median, O2Sat_min, Temp_median, Temp_max, SBP_median, SBP_min, MAP_median, MAP_min, DBP_median, DBP_min, Resp_median, Resp_max
- **BGA/respiratory:** FiO2_median, BaseExcess_median, HCO3_median, pH_median, PaCO2_median, SaO2_median
- **counts:** HCO3_count, PaCO2_count, SaO2_count, Lactate_count
- **labs:** BUN_median, Creatinine_median, Glucose_median, Lactate_median, Potassium_median, Bilirubin_total_median, Hgb_median, WBC_median, Platelets_median
- **spike labs:** Glucose_max
- **categoric/target:** Gender_first, ICU_Unit, SepsisLabel

# Modeling & Hiperparameter Tuning

| Metric | What it measures | Choosing the right metric for this project | |
|---|---|---|---|
| **Precision** | Out of **all patients** flagged as **sepsis by the model**, how many **are truly sepsis** | Can be high even when many true sepsis cases are missed **(low recall)** | ✗ |
| **Recall** | Out of **all true sepsis patients,** how many are **correctly identified by the model** | Can be high if the model labels almost everyone as sepsis (**very high false alarms**) | ✗ |
| **F1-score** | Balance between precision and recall **at a single threshold** | Only reflects **performance at one threshold**; does not show behavior at other thresholds | ✗ |
| **AUPRC** | Precision–recall **pattern across all thresholds**, focusing on the minority class | **More appropriate for imbalanced data** and for evaluating performance on sepsis (minority class) | ✓ |

# Modeling & Hiperparameter Tuning

- **AUPRC is used** for modelling because **sepsis is a rare event (~7%)**, so metrics like accuracy or AUROC can look high even when the model fails to capture the minority class.
- **AUPRC focuses on the positive class** by summarizing the precision–recall trade-off across thresholds, matching the goal of early sepsis detection with controlled false alarms.
- Sepsis prevalence is ~7%, so the **baseline AUPRC equals 0.07.**
- AUPRC = 0.07 indicates random guessing with no predictive value.
- **AUPRC > 0.07** shows the model **performs better than random at identifying sepsis.**

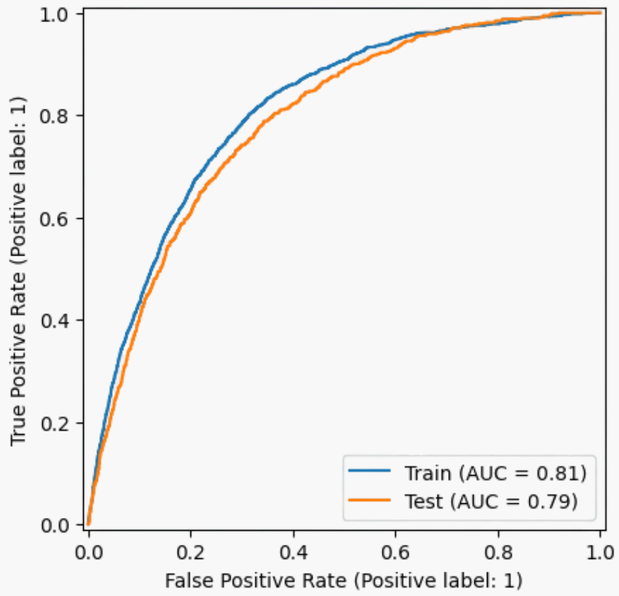
# Modeling & Hiperparameter Tuning

| MODEL | AUPRC DEFAULT | AUPRC TUNING |
|---|---|---|
| Decision Tree | 0,094 | 0.156 |
| Random Forest | 0.205 | 0.244 (Overfitting) |
| XGBOOST | 0.228 (Overfitting) | 0.212 |

# Model Evaluation

```
AUPRC Score (Train): 0.2425179642789735
Best threshold from val: 0.103905804
Classification Report Xgboost (Train):
              precision    recall  f1-score   support

           0       0.96      0.84      0.90     23955
           1       0.22      0.58      0.32      1849

    accuracy                           0.83     25804
   macro avg       0.59      0.71      0.61     25804
weighted avg       0.91      0.83      0.86     25804
```

```
AUPRC Score (Test): 0.2122085792831916
Classification Report Xgboost (Test):
              precision    recall  f1-score   support

           0       0.96      0.84      0.90      7484
           1       0.21      0.55      0.30       581

    accuracy                           0.82      8065
   macro avg       0.59      0.70      0.60      8065
weighted avg       0.91      0.82      0.85      8065
```




- Test **AUPRC = 0.212** vs **baseline 0.07** shows (sepsis prevalence) shows a **clear early-warning signal:** the model is roughly 3x better than random at ranking patients by sepsis risk. **ROC-AUC = 0.79** (train 0.81, test 0.79) indicates **good overall discrimination** using only the first 24 hours of ICU data.
- With the tuned **threshold 0.104**, the model achieves **recall ≈ 0.55** for sepsis (captures **more than half of sepsis cases early**) and **precision ≈ 0.21** (around **1 in 5 alerts is a true sepsis case**), which is expected for a sensitive early-warning tool.
- The **gap between train and test is small** (AUPRC 0.24 → 0.21; ROC-AUC 0.81 → 0.79), suggesting only mild overfitting and **reasonable generalisation to unseen ICU patients**.

# Model Evaluation

## Confusion Matrix

| | 0 | 1 |
|---|---|---|
| 0 | 6277 | 1207 |
| 1 | 260 | 321 |

- **TN = 6277**: non-sepsis correctly identified.
- **FP = 1207**: non-sepsis incorrectly flagged as sepsis (false alarms).
- **FN = 260**: sepsis missed by the model.
- **TP = 321:** sepsis correctly detected.

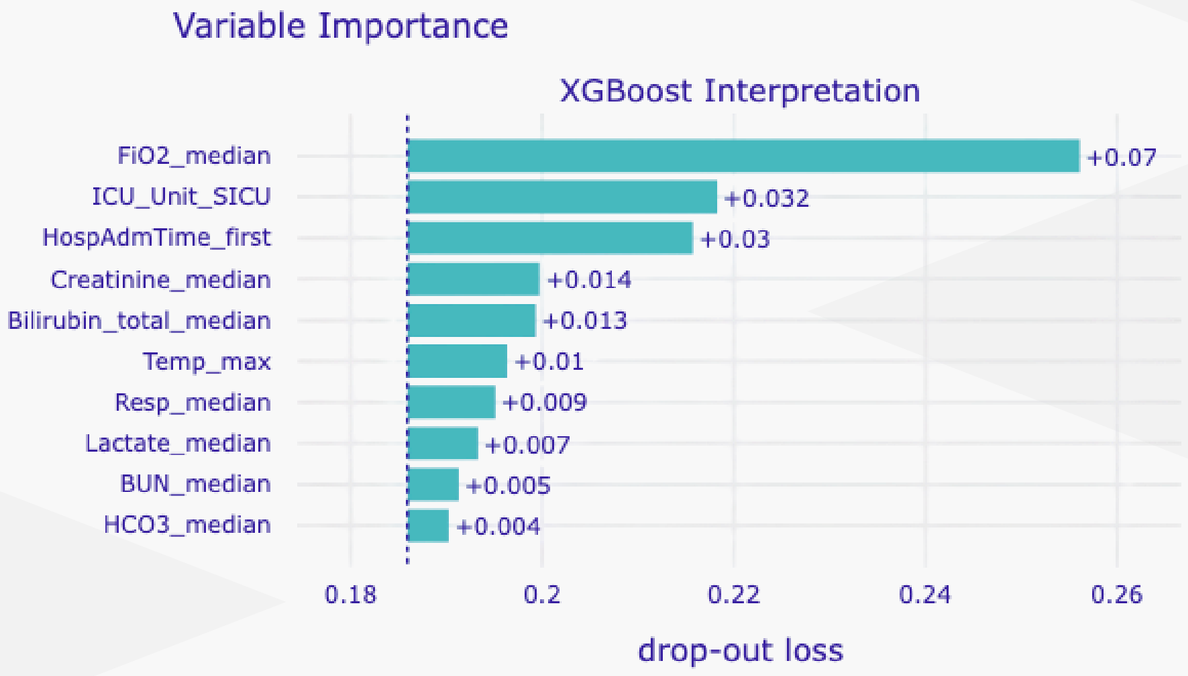
- The **model detects about 55% of sepsis cases** in the first 24 hours, which is consistent with its **role as an early-warning tool rather than a diagnostic test.**
- **False alarms are still high** (878 FP vs 294 TP) and will **require clinical triage.**
- **Missed sepsis** remains substantial (**260 FN**). The model should not be used as a stand-alone decision maker.

# Recommendations for the hospital

- Use **the model as a clinical early-warning tool** that requires clinician confirmation, not as an automatic diagnosis system.
- **Create a clear response pathway** : **model alert → rapid nurse/physician review → initiate sepsis bundle if clinically indicated**.
- **Start with a pilot phase** before full rollout **to assess impact on time-to-antibiotics, false-alarm burden, and patient outcomes.**
- **Monitor and adjust regularly.** Track TP/FP/FN/TN, recalibrate threshold or retrain model if needed.
- **Train ICU staff.** Provide brief guidance on how to interpret the risk score and integrate alerts into existing sepsis workflow.

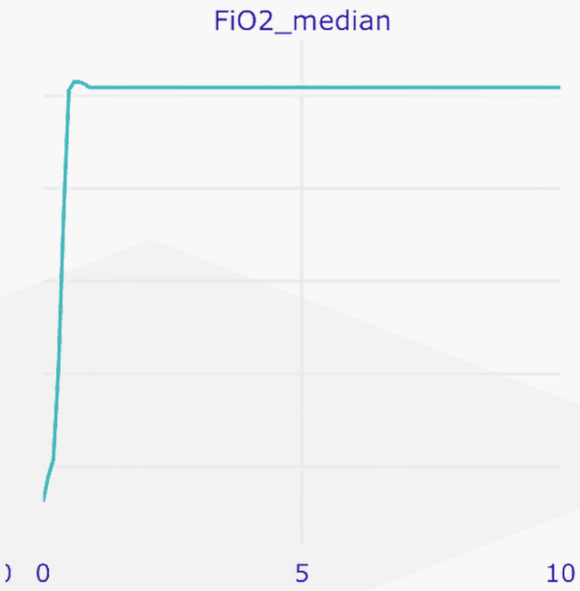
***The model prioritizes early detection and screening; providing risk signals to trigger faster clinical evaluation rather than final diagnosis.***

# Features Importance



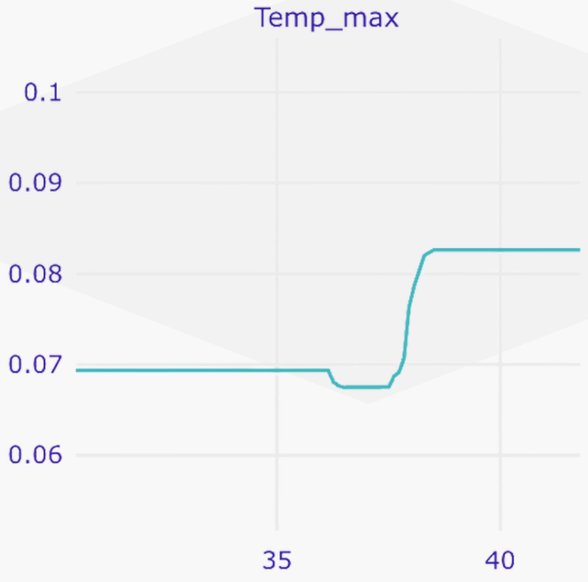

**FiO2_median** is the **most dominant feature** determining model accuracy, far exceeding others, which indicates that respiratory status is the key driver of prediction. Notably, operational factors like **ICU unit type** and **admission time** proved more influential than biological markers for kidney or liver function.

# Partial Dependence




**$FiO_2$_median > 0.5** → Predicted sepsis risk rises sharply when the patient needs >50% oxygen.

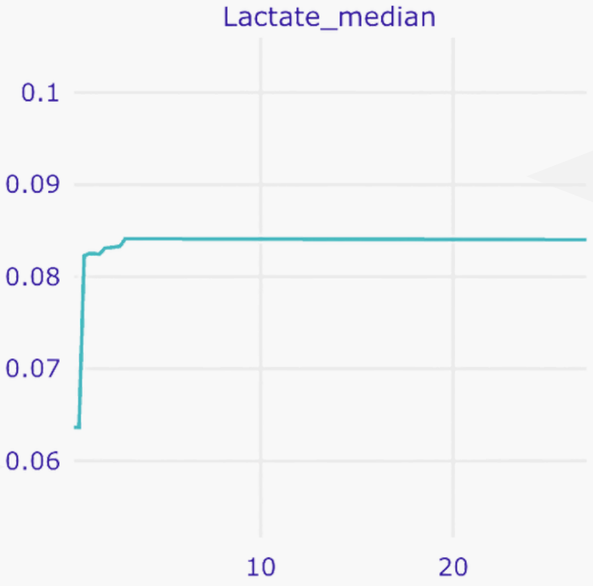
**Resp_median ≥ 24/min →** Higher respiratory rate is associated with higher sepsis risk.



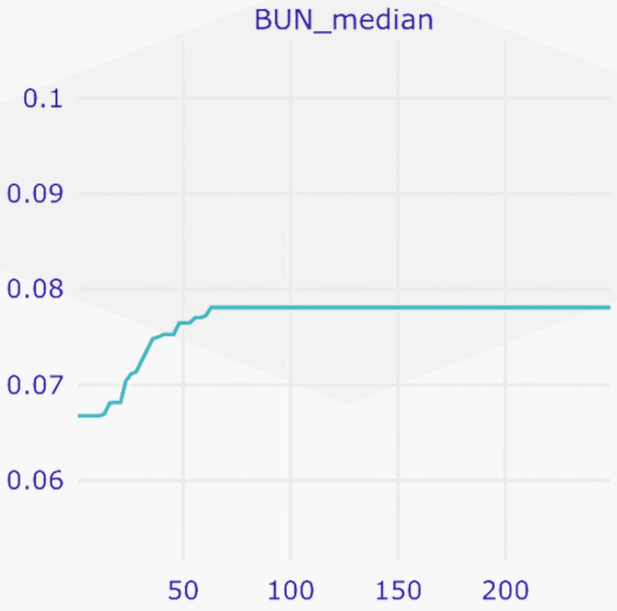

**Temp_max ≥ 38°C** → Fever range shows an increase in predicted sepsis probability.

# Partial Dependence

**Creatinine_median ≥ 2 mg/dL →**
Worsening kidney function corresponds to higher sepsis risk.

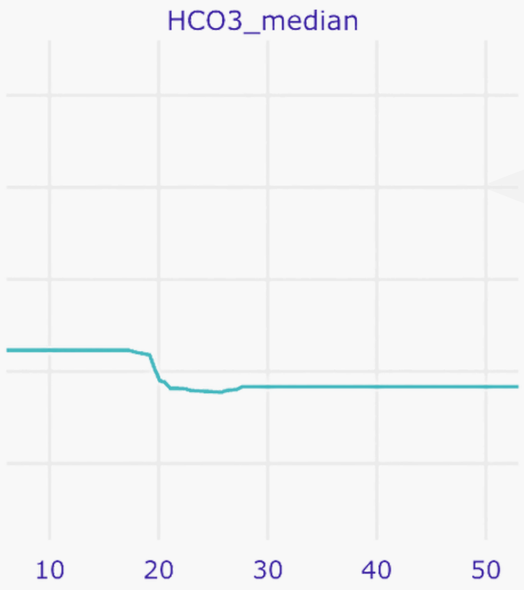



**Lactate_median ≥ 2 mmol/L →**
Elevated lactate is linked with a clear jump in sepsis risk.



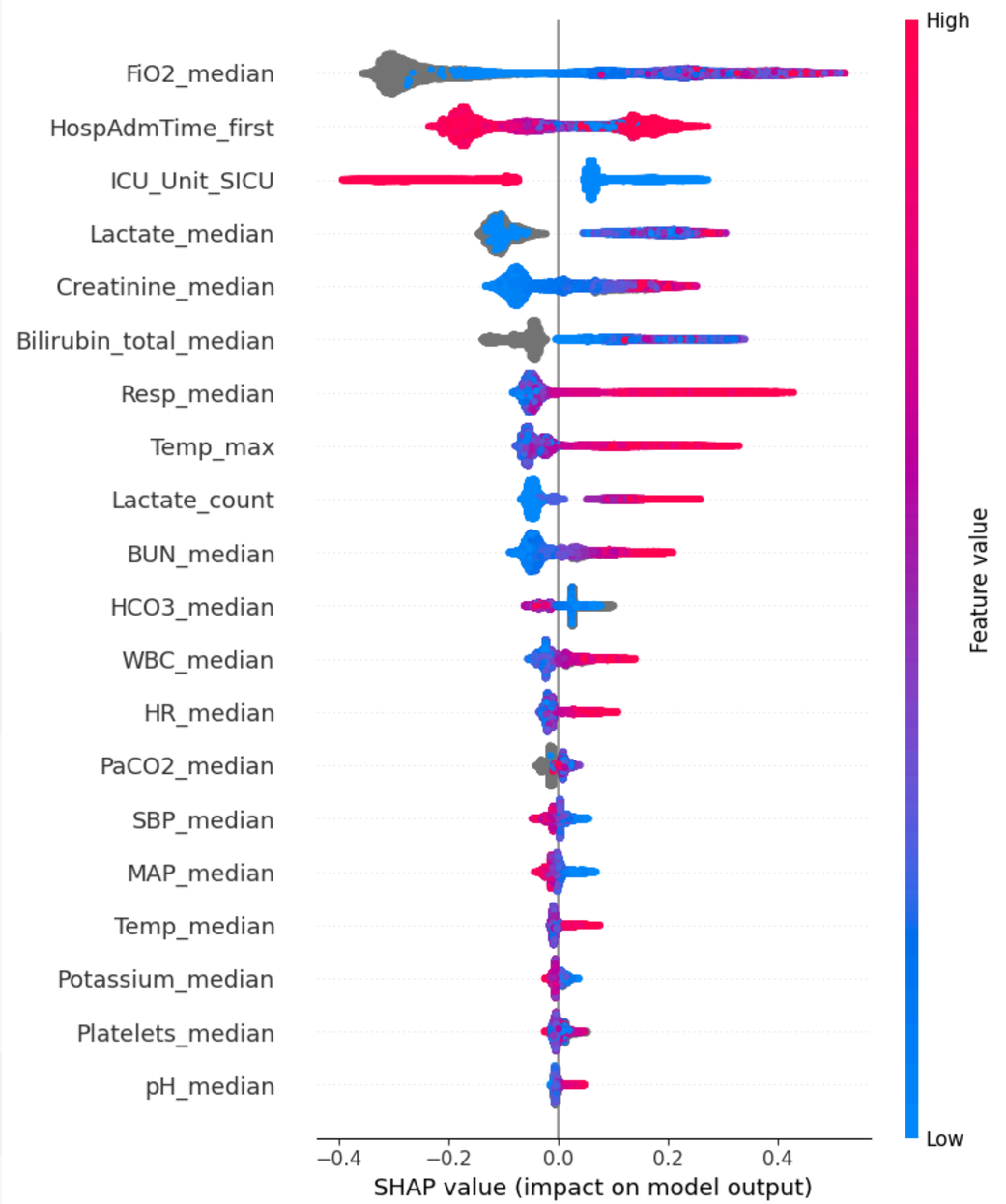

**BUN_median ≥ 60 mg/dL →**
Marked azotemia is associated with increased predicted risk.

# Partial Dependence




**$HCO_3$_median ≤ 19 mEq/L** →
Metabolic acidosis (low $HCO_3^-$) is followed by higher sepsis risk.

# Shap Model




- **The most influential features** are FiO2_median, HospAdmTime_first, ICU_Unit_SICU, Lactate_median, Creatinine_median, and Bilirubin_total_median.
- **Higher values of** $FiO_2$, lactate, creatinine, respiratory rate and temperature show positive SHAP values, indicating **higher predicted sepsis risk.**
- Being admitted to **SICU** (ICU_Unit_SICU = 1) is also **associated with increased model-predicted risk** compared with other ICU units.

# Deployment

https://early-sepsis-prediction.streamlit.app/

# Deployment

Add this observation

**Recorded observations (all times):**

| | time | HR | O2Sat | Temp | SBP | MAP | DBP | Resp | FiO2 | BaseExcess | HCO3 | pH | PaCO2 | SaO2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2025-11-28 12:00:00 | 120 | 92 | 37.5 | 100 | 80 | 70 | 26 | 0.6 | None | 20 | 7.4 | 28 | 93 |
| 1 | 2025-11-28 15:00:00 | 144 | 90 | 37.5 | 90 | 70 | 60 | 28 | 0.75 | None | 19 | 7.3 | 35 | 89 |
| 2 | 2025-11-28 16:15:00 | 132 | 94 | 38 | 110 | 80 | 65 | 30 | 0.7 | None | 19 | 7.29 | 36 | 91 |

Compute 24h Features & Predict

## Aggregated 24h Features (sent to model)

| | HR_median | HR_max | O2Sat_median | O2Sat_min | Temp_median | Temp_max | SBP_median | SBP_min | MAP_median | MAP_ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 138 | 144 | 92 | 90 | 37.75 | 38 | 100 | 90 | 75 | |

## Prediction Result

**Sepsis probability (model output):** 0.185

High risk of sepsis (Positive flag)

Use as an early warning signal; clinical confirmation is still required.

https://early-sepsis-prediction.streamlit.app/

# Conclusion

This project shows that **sepsis risk** in ICU patients **can be predicted from routine** data in the first 24 hours using an XGBoost model with reasonable discrimination (AUPRC clearly above baseline). Higher $FiO_2$, SICU admission, longer pre-ICU hospital stay, abnormal vital signs, metabolic acidosis, and organ-dysfunction markers (lactate, creatinine, bilirubin, BUN) emerge as the strongest early risk signals.

## Future directions

- **For hospitals:** Implement the model as an **early-warning decision** support tool (not an automatic diagnosis) and define clear **clinical response pathways** for high-risk alerts.
- **For the model:** It can be improved by adding richer information **(key medication/procedure data**) so that predictions become more accurate and clinically meaningful. The model can also be **integrated with the hospital system** so clinicians can give feedback on alerts, and this feedback is used to continuously refine the model.

# Tools & Libraries Used

## Language & Environment

python
CO
git
GitHub

## Core Data & Scientific Stack

pandas
NumPy
SciPy
scikit learn

## Modeling, Tuning & Explainability

XGBoost
OPTUNA
DALEX

## Visualization & App

seaborn
matplotlib
Streamlit

# Thank You