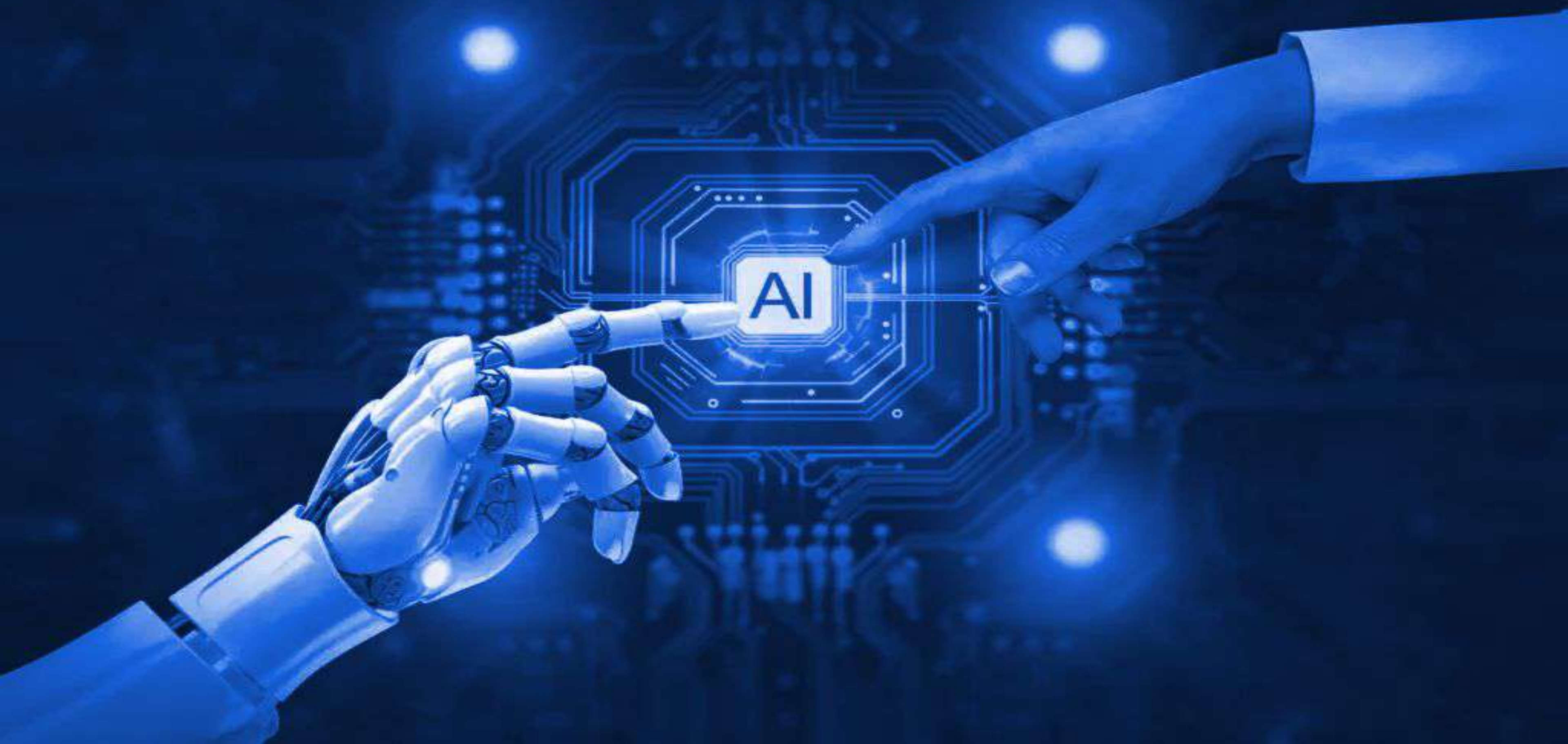
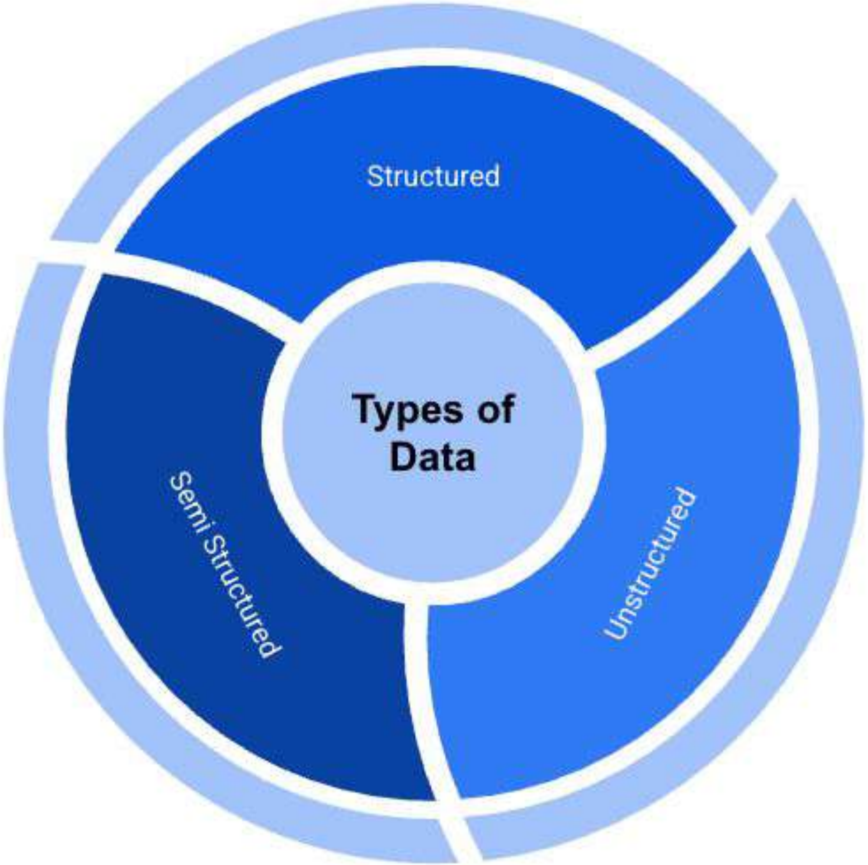
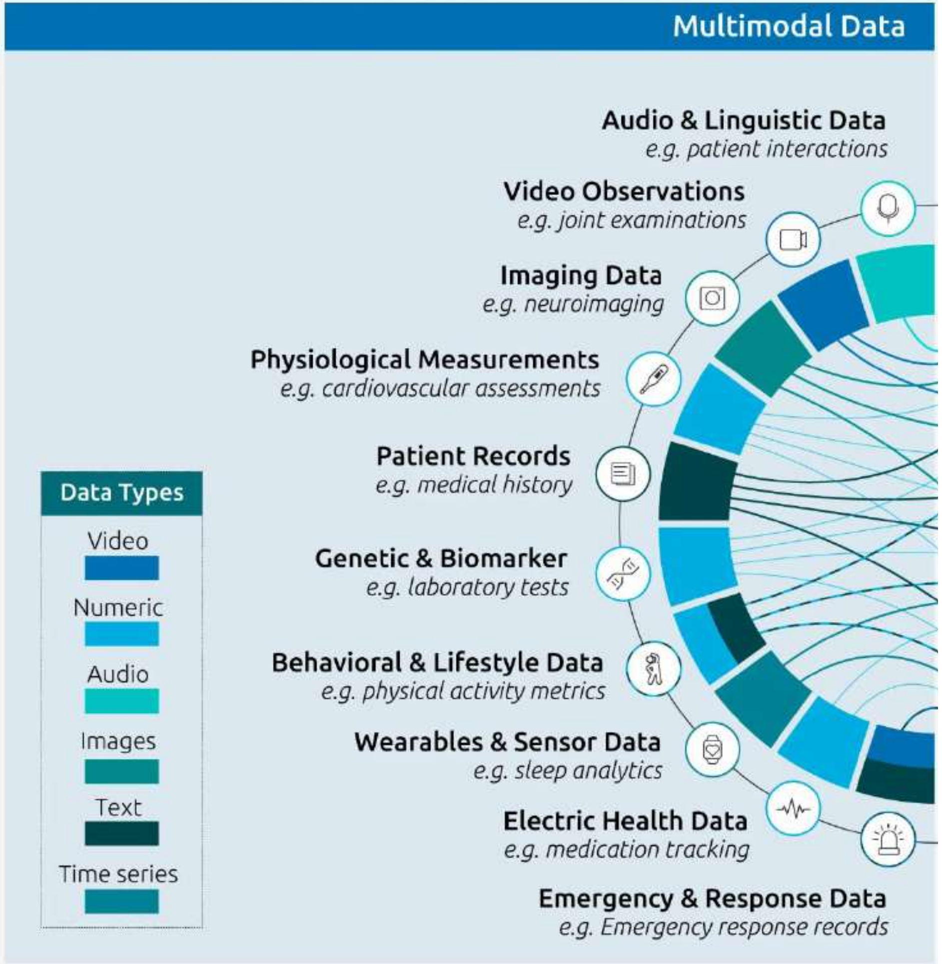
AI for Multimodal Healthcare Data



- Why AI in healthcare ?

- Types of Healthcare Data




Multimodal Data
Data Types
Video
Numeric
Audio
Images
Text
Time series
Audio & Linguistic Data
e.g. patient interactions
Video Observations
e.g. joint examinations
Imaging Data
e.g. neuroimaging
Physiological Measurements
e.g. cardiovascular assessments
Patient Records
e.g. medical history
Genetic & Biomarker
e.g. laboratory tests
Behavioral & Lifestyle Data
e.g. physical activity metrics
Wearables & Sensor Data
e.g. sleep analytics
Electric Health Data
e.g. medication tracking
Emergency & Response Data
e.g. Emergency response records

Case Study – ShockModes

Shock: What is it?

- A life-threatening condition where tissues don't get enough oxygen and nutrients. Leads to organ failure if untreated.

Causes of Shock

- Severe blood or fluid loss; heart fails to pump effectively; widespread infection causes blood vessel dilation + leakage.

Why Treating Early Matters

- Mortality Rates : 30 to 40%

Shock Index (SI)

- Heart Rate (HR)/Systolic Blood Pressure (SBP)
- Abnormal: SI ≥ 0.7 → higher risk of circulatory collapse

ShockModes Goal

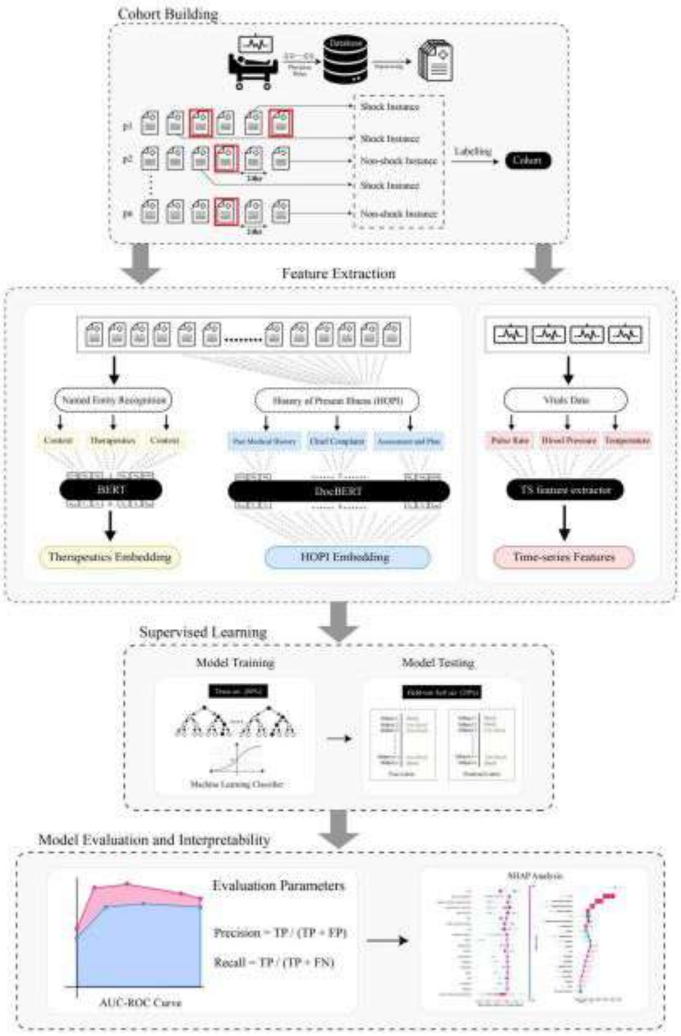
- Build an Early Warning System.

- Predict abnormal SI 24 hours in advance.

- Use both vitals + physician notes.

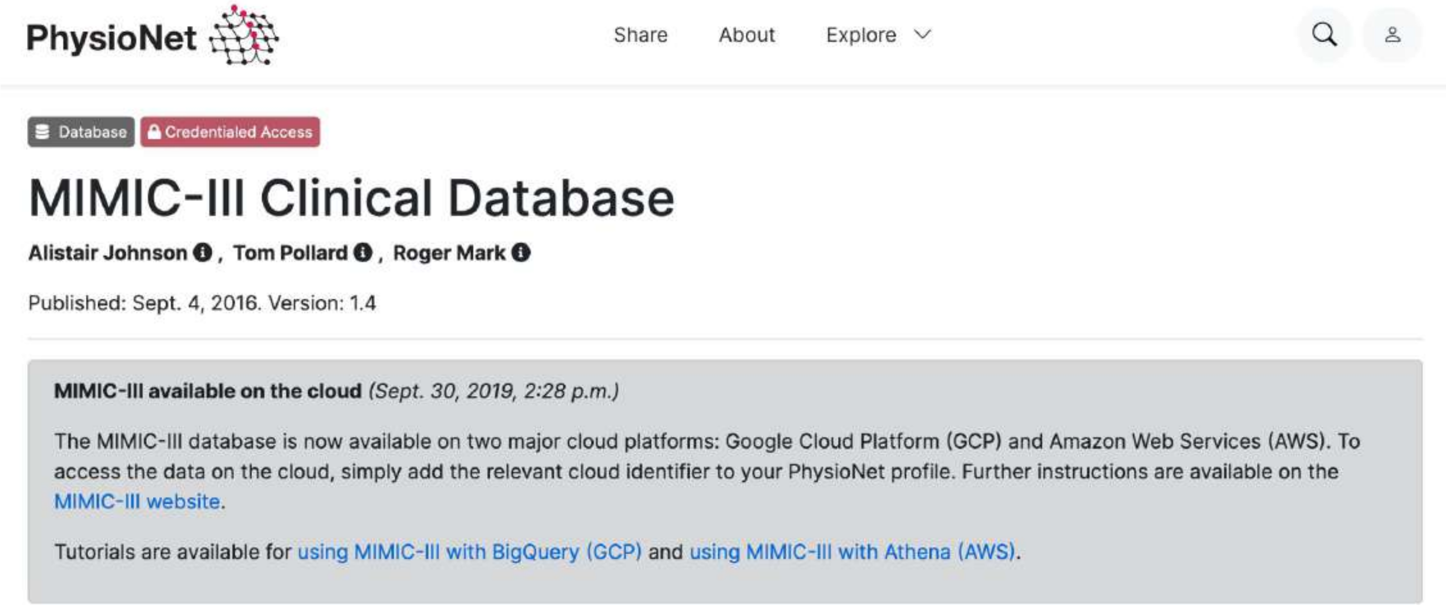
Workflow

- Collect vitals + notes (MIMIC-III dataset).
- Clean & preprocess data.
- Extract features:
  - Vitals → time-series patterns.
  - Notes → keywords, embeddings.
- Fuse features.
- Train ML models (Random Forest, Boosting).
- Output: Risk prediction for abnormal SI.

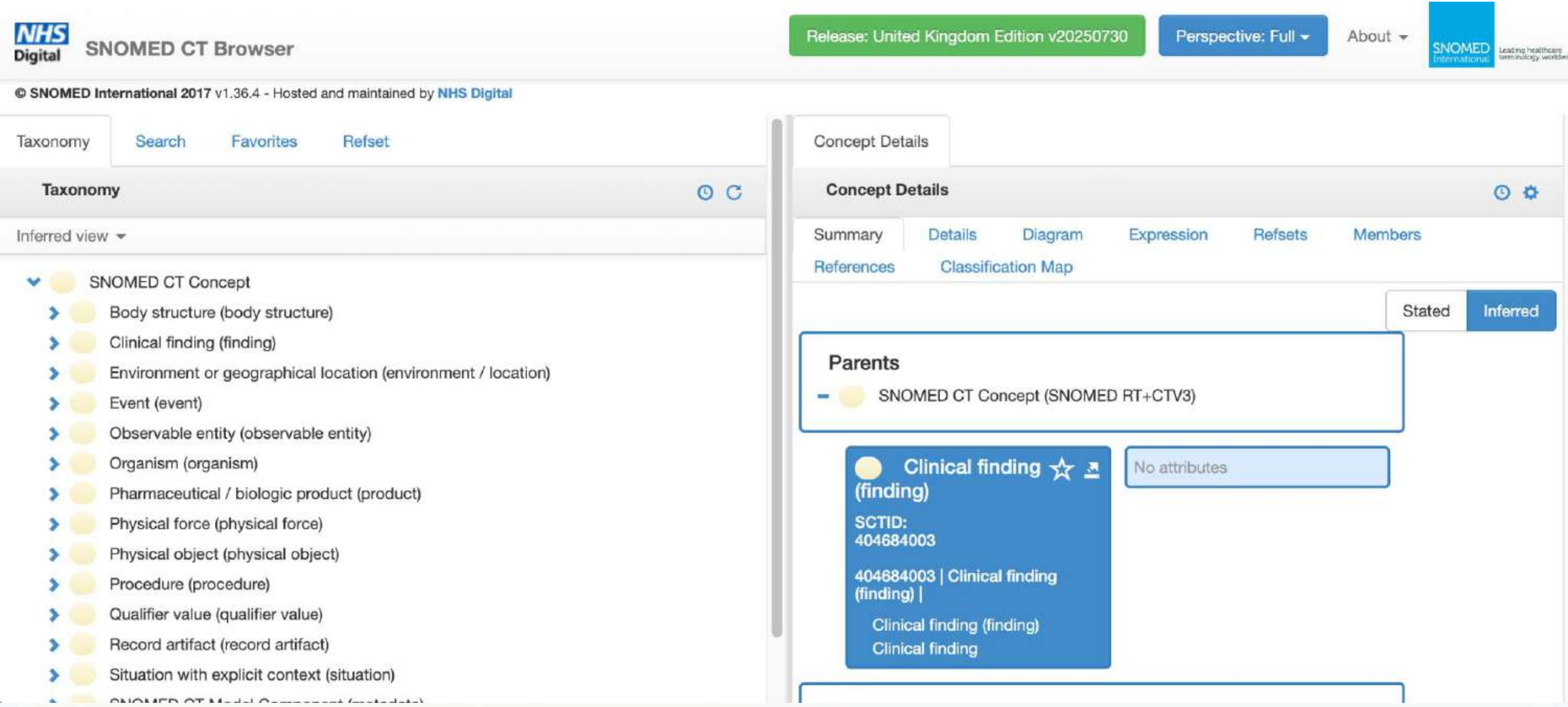


Dataset

- Used MIMIC-III database with ICU data from 17,294 ICU-stays.
- Curated 24-hour patient encounter cohorts combining Noteevents and vital signs.
- Extracted key vitals: HR, SBP, RR, and SpO_2 for each 24-hour window.



Features from Notes



Features from Vitals

- Extracted 3117 time-series features using tsfresh.

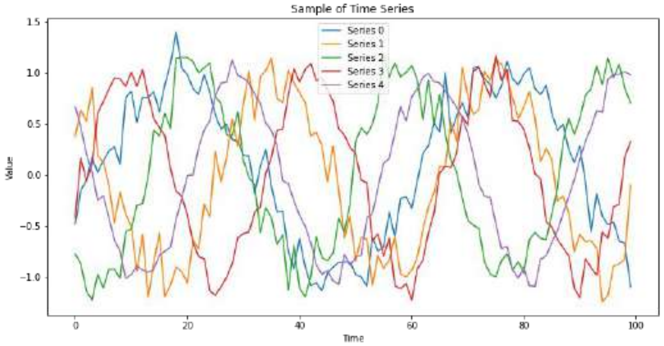
- Examples:

  Autocorrelation (patterns)

  Fourier transforms (signal changes)

  Entropy (randomness in signals)

- tsfresh converts raw time-series into thousands of features (mean, variance, entropy, Fourier coefficients, etc.).
- Extracted features can be fed into ML models for tasks like classification, regression, clustering, or forecasting.

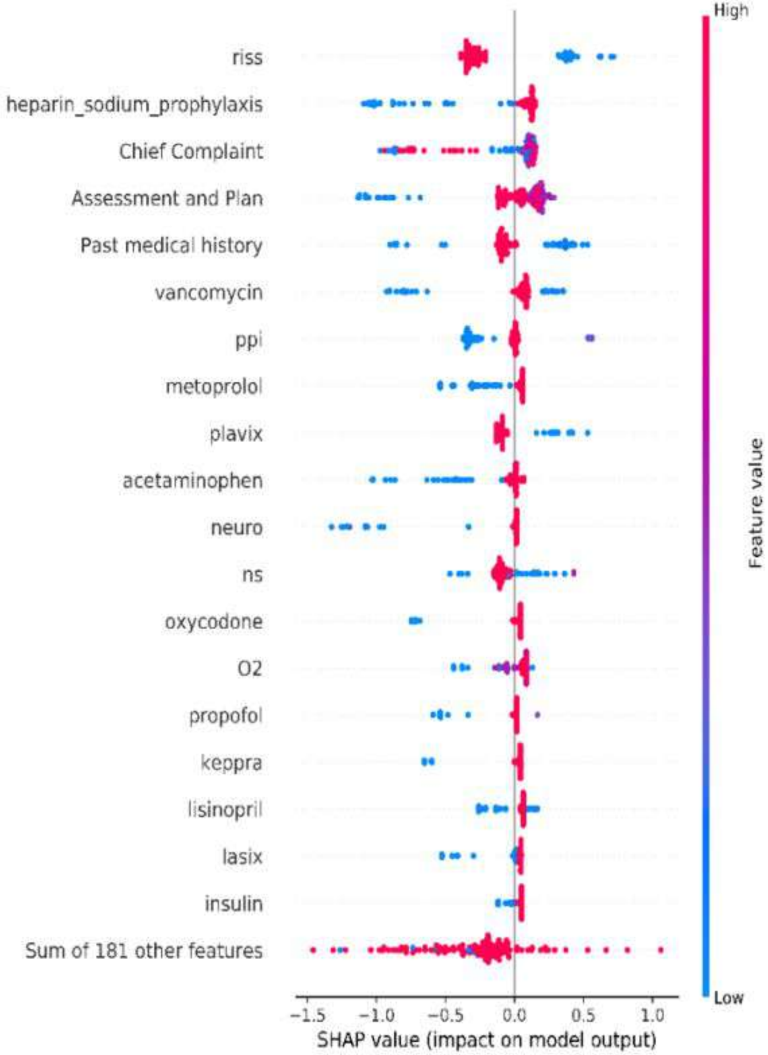


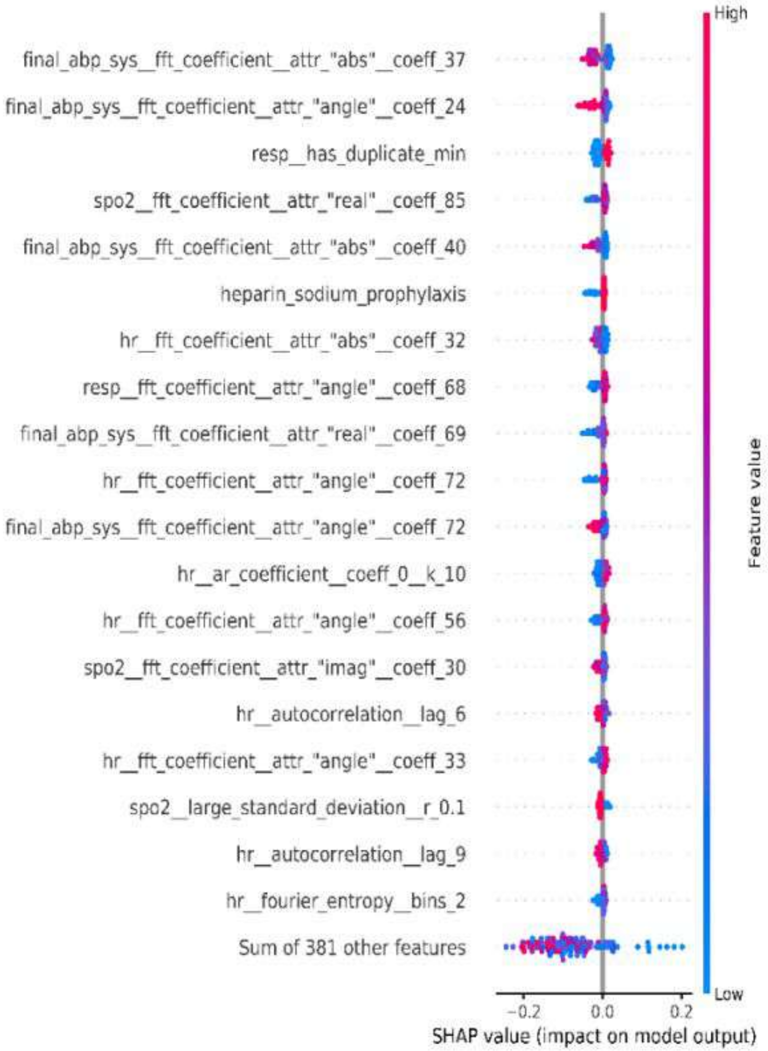
Model Development and Validation

- Input features: vitals (tsfresh), embeddings of therapeutics, and HOPI.
- Feature reduction: Used Extra Trees classifier (scikit-learn) to select important features.
- Imbalanced data: Handled using SMOTE oversampling (scikit-learn).
- Models used: Logistic Regression, Random Forest, GradientBoost, AdaBoost, XGBoost.
- Evaluation: Performed with bootstrap sampling (100 iterations); key metrics were AUC-ROC and F1-score.

SHAP Analysis for Interpretability

- SHAP (Shapley Additive Explanations): Uses Shapley values to calculate feature importance and explain model outputs.
- Interpretability method: Prior probability from training data is compared with model output probability to explain predictions.
- Local explanations: Patient-specific predictions visualized using waterfall plots, showing which features pushed toward normal/abnormal SI.
- Global explanations: Overall feature importance visualized with bar plots and beeswarm plots.


riss
heparin_sodium_prophylaxis
Chief Complaint
Assessment and Plan
Past medical history
vancomycin
ppi
metoprolol
plavix
acetaminophen
neuro
ns
oxycodone
O2
propofol
keppra
lisinopril
lasix
insulin
Sum of 181 other features
−1.5 −1.0 −0.5 0.0 0.5 1.0
SHAP value (impact on model output)
High
Feature value
Low


final_abp_sys__fft_coefficient__attr_"abs"__coeff_37
final_abp_sys__fft_coefficient__attr_"angle"__coeff_24
resp__has_duplicate_min
spo2__fft_coefficient__attr_"real"__coeff_85
final_abp_sys__fft_coefficient__attr_"abs"__coeff_40
heparin_sodium_prophylaxis
hr__fft_coefficient__attr_"abs"__coeff_32
resp__fft_coefficient__attr_"angle"__coeff_68
final_abp_sys__fft_coefficient__attr_"real"__coeff_69
hr__fft_coefficient__attr_"angle"__coeff_72
final_abp_sys__fft_coefficient__attr_"angle"__coeff_72
hr__ar_coefficient__coeff_0__k_10
hr__fft_coefficient__attr_"angle"__coeff_56
spo2__fft_coefficient__attr_"imag"__coeff_30
hr__autocorrelation__lag_6
hr__fft_coefficient__attr_"angle"__coeff_33
spo2__large_standard_deviation__r_0.1
hr__autocorrelation__lag_9
hr__fourier_entropy__bins_2
Sum of 381 other features
−0.2 0.0 0.2
SHAP value (impact on model output)
High
Feature value
Low