Early prediction of sepsis from time-series clinical data

Anna Rechkalova

Supervisor: Dmytro Kuzmenko,

Sr. lecturer, Dep. of Multimedia Systems

Research Relevance & Current Approaches

- **Sepsis kills 21.4 million people annually**
- **Over 31% of global mortality**
- **Every hour of delayed treatment increases mortality risk by 7%**
- **Ukrainian context: wartime trauma and wound infections significantly increase sepsis incidence, while simultaneously overloading critical care infrastructure**

Current Approaches: SOFA & qSOFA

Research Objectives

Goal: predict sepsis 6 hours before clinical onset.

- **Which modelling paradigm is more effective for early sepsis prediction from ICU time series: classical gradient boosting or deep learning with learned temporal representations?**
- **Does a hybrid architecture combining both paradigms outperform each component individually?**
- **Do local post-hoc explainability methods produce clinically meaningful attributions for ICU time series predictions?**

Dataset & Exploratory Data Analysis

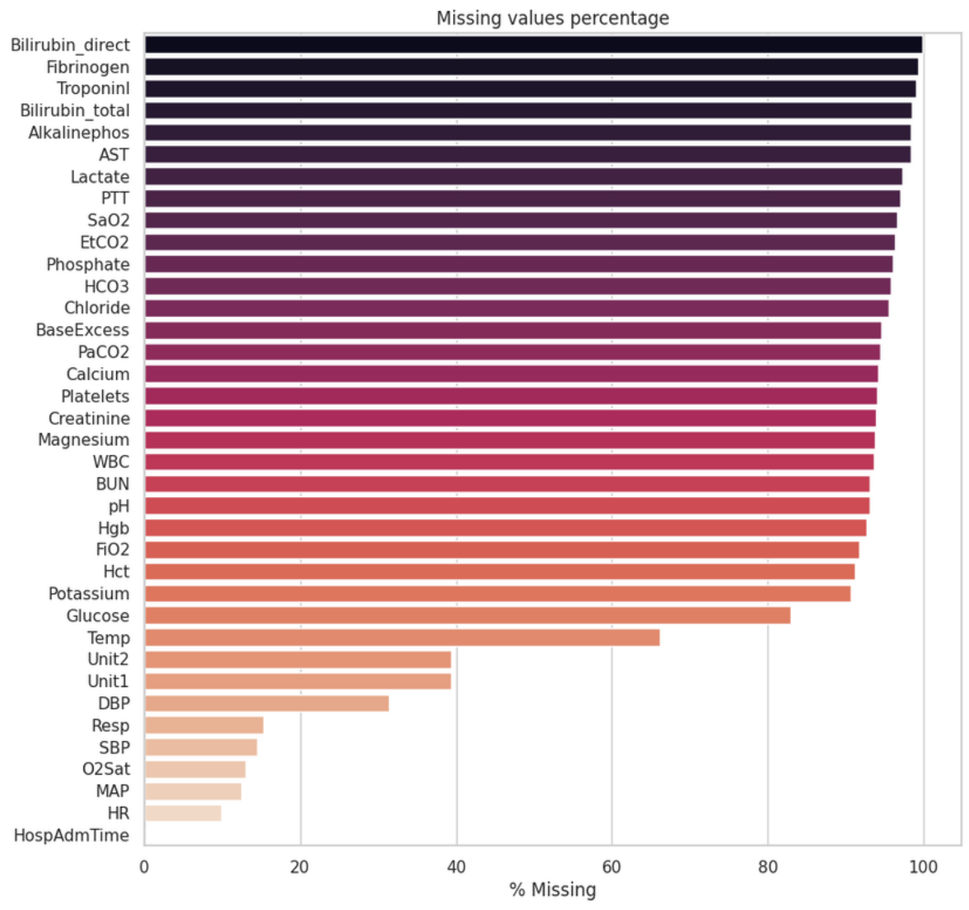


Figure: Missing values percentage for all features

- **PhysioNet/CinC Challenge 2019 [2]**
- **40,336 ICU patients (combined sets A and B; 1,552,210 records)**
- **40 hourly clinical variables:**
  - **8 Vital Signs (HR, SBP, MAP, Temp, O2Sat…)**
  - **26 Lab Values (Lactate, Creatinine, FiO2…)**
  - **6 Demographics (Age, Gender, ICULOS…)**
- **Sepsis-3 labels shifted 6h forward**
- **Class imbalance (No sepsis: 92.7%, sepsis: 7.3%)**

Key EDA finding: Missing Not At Random (MNAR)

Missingness itself carries diagnostic information

Data Preprocessing & Engineering

Forward-fill + Missing Indicators

Fill gaps with last known value

Add binary mask: 1 = measured, 0 = filled

+36 missing indicator features

Manual Clinical Features

Shock Index = HR / SBP

BUN/CR ratio

+2 features

Sliding Window Aggregates

Vital signs: min, max, mean, std in 2h, 3h, 6h, 12h windows

Lab values: mean + frequency in 6h, 12h windows

+162 engineered features

For gradient-boosting method

240 features were derived after the data processing

Models

- **Gradient-boosting approach: LightGBM (LGBM)**
- **Deep learning approach: Temporal Convolution Network (TCN)**
- **Hybrid: TCN as Feature Extractor => LightGBM**

Dataset split: 75% (train), 15% (validation), and 10% (test).

Metrics: AUPRC and Utility score.

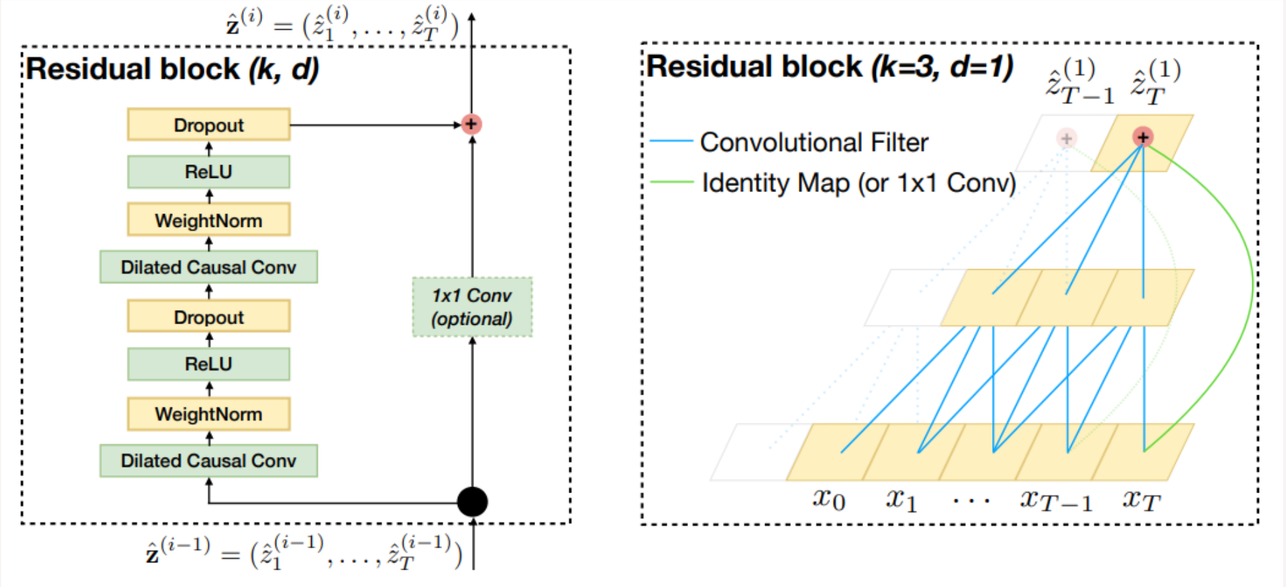


Figure: TCN residual block & An example of residual connection in a TCN

Experiments

Table: Comparison of feature selection strategies and hyperparameter optimization on the validation set. Bold values indicate the final configuration selected for each model class

Model	Feature Set	Number of Features	Thr	AUPRC	Utility
LightGBM	All Features (Baseline)	240	0.20	0.1168	0.4309
	Cumulative sum 99%	122	0.20	0.1231	0.4323
	Cumulative sum 95%	75	0.19	0.1183	0.4273
	Gain above mean	28	0.22	0.1120	0.4230
	Hyperparameter optimization	**122**	**0.21**	**0.1371**	**0.4614**
TCN	All Features (Baseline)	240	0.16	0.1185	0.3806
	Original + Missing	76	0.18	0.1316	0.4263
	Pos. Importance (from 240)	145	0.24	0.1294	0.3926
	Pos. Importance (from 76)	**54**	**0.22**	**0.1285**	**0.4369**
	Hyperparameter optimization	54	0.24	0.1316	0.4266
Hybrid	TCN Embeddings Only	128	0.27	0.1091	0.4213
	Embeddings + Feats used for TCN	182	0.22	0.1073	0.4298
	Embeddings + Original Features	168	0.20	0.1141	0.4252
	Hyperparameter optimization	**168**	**0.23**	**0.1278**	**0.4473**

Results

Table: Model performance comparison

Model	Val AUPRC	Val Utility	Test AUPRC	Test Utility
LightGBM	**0.137**	**0.461**	**0.116**	**0.381**
TCN	0.129	0.437	0.105	0.374
Hybrid	0.127	0.433	0.091	0.365

Table: Clinical utility scores for the teams with the five highest scores on the full test set from hospital systems A, B, and C

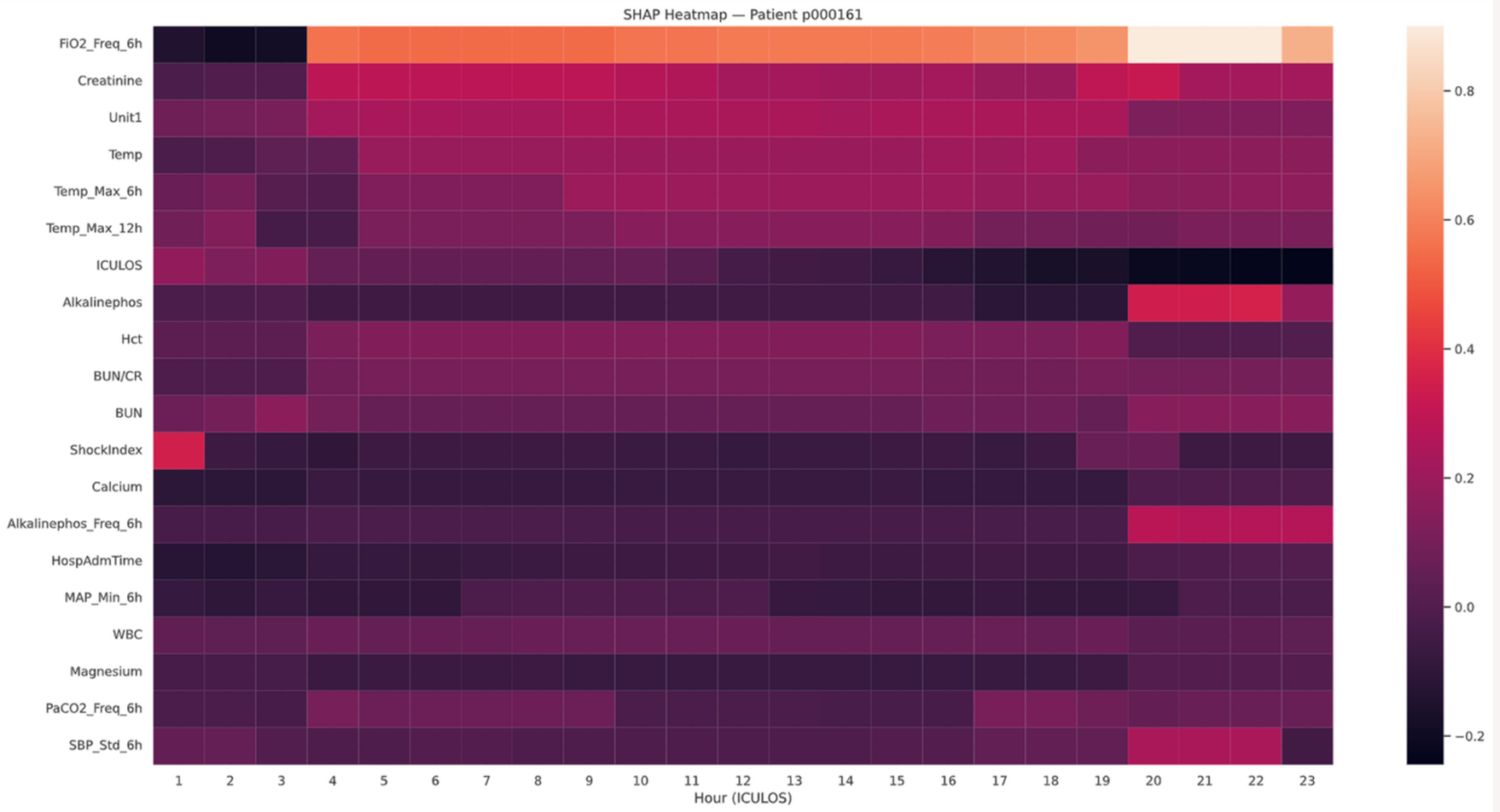
Rank	Team	Final Score	Score A	Score B	Score C
1	James Morrill, Andrey Kormilitzin, Alejo Nevado-Holgado, Sumanth Swaminathan, Sam Howison, Terry Lyons	0.360	0.433	0.434	-0.123
2	John Anda Du, Nadi Sadr, Philip de Chazal	0.345	0.409	0.396	-0.042
3	Morteza Zabihi, Serkan Kiranyaz, Moncef Gabbouj	0.339	0.422	0.395	-0.146
4	Xiang Li, Yanni Kang, Xiaoyu Jia, Junmei Wang, Guotong Xie	0.337	0.420	0.401	-0.156
5	Janmajay Singh, Kentaro Oshiro, Raghava Krishnan, Masahiro Sato, Tomoko Ohkuma, Noriji Kato	0.337	0.401	0.407	-0.094

Explainability

- **Local Ad-hoc methods:**
  - **LGBM’s approach: SHAP (SHapley Additive exPlanations)**
  - **TCN’s approach: Windowed Temporal Saliency Rescaling**

- **Task: Check consistency with the feature importance evaluated for the feature selection stage**

SHAP Heatmap



- **Raw & Clinical indices: 13 (including BUN/CR and ShockIndex)**
- **Window = 4**
- **Frequency = 3**

Figure: SHAP Heatmap – Patient p000161. Top-20 features

SHAP Bar chart

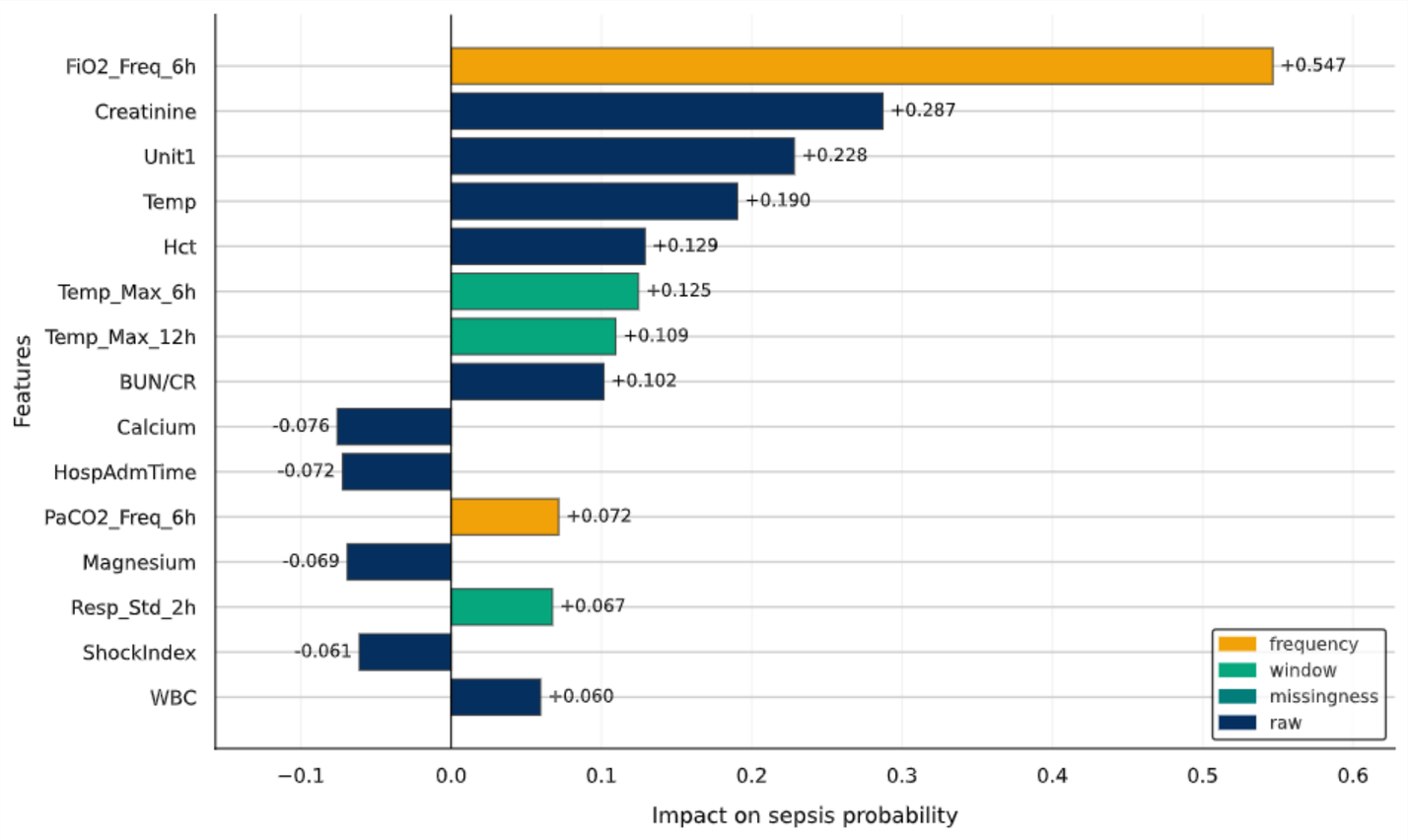


Figure: SHAP Bar chart – Patient p000161 (Hour 6). Top-15 features

FiO2_Freq_6h dominates,

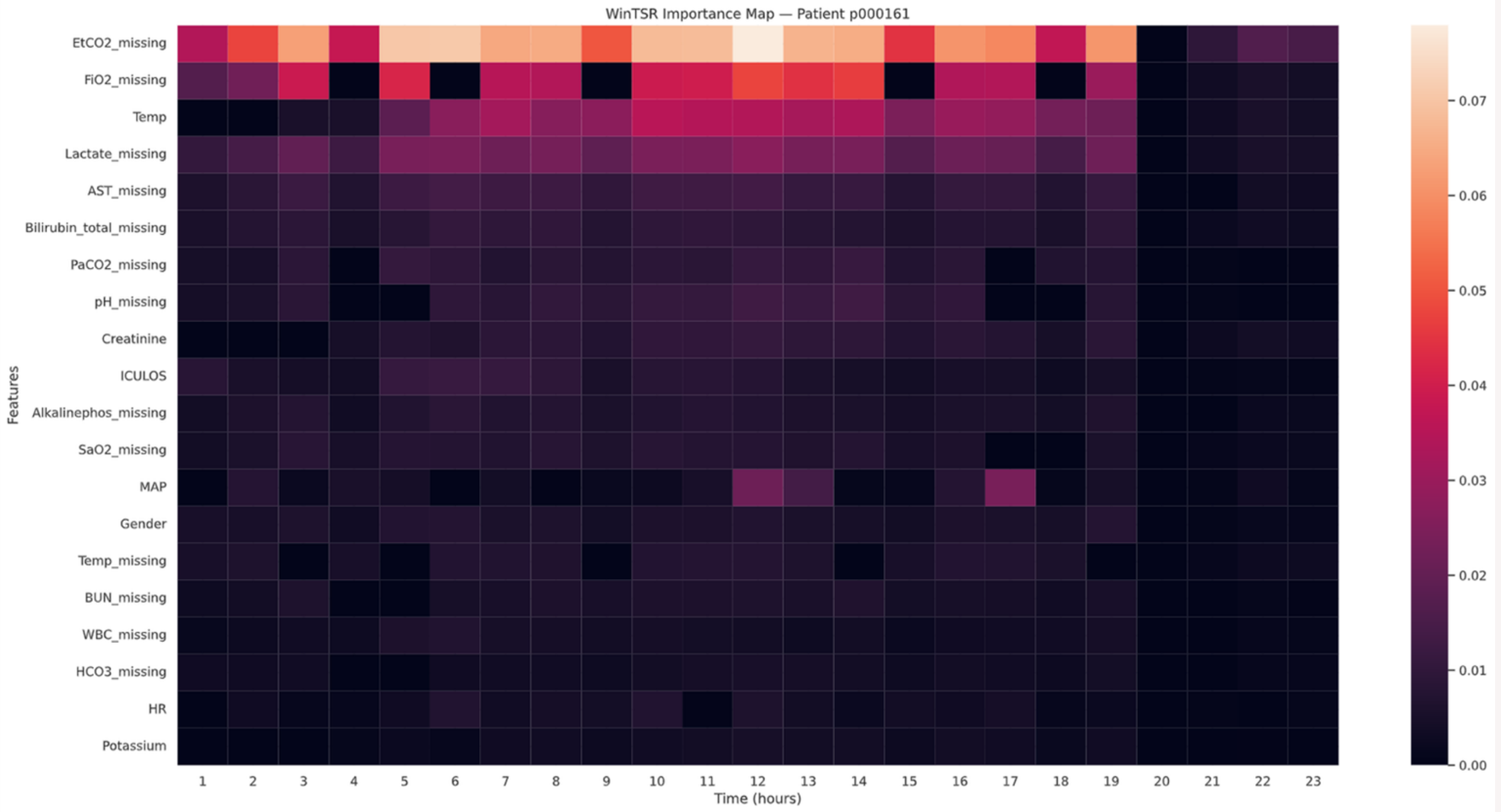
consistent with MNAR hypothesis.

= FiO2 is ordered more frequently

when respiratory support is

needed.

WinTSR Heatmap



- **Missingness = 13**
- **Raw = 7**

= TCN learns temporal patterns directly from the sequence.

Figure: WinTSR Heatmap – Patient p000161. Top-20 features

WinTSR Bar chart

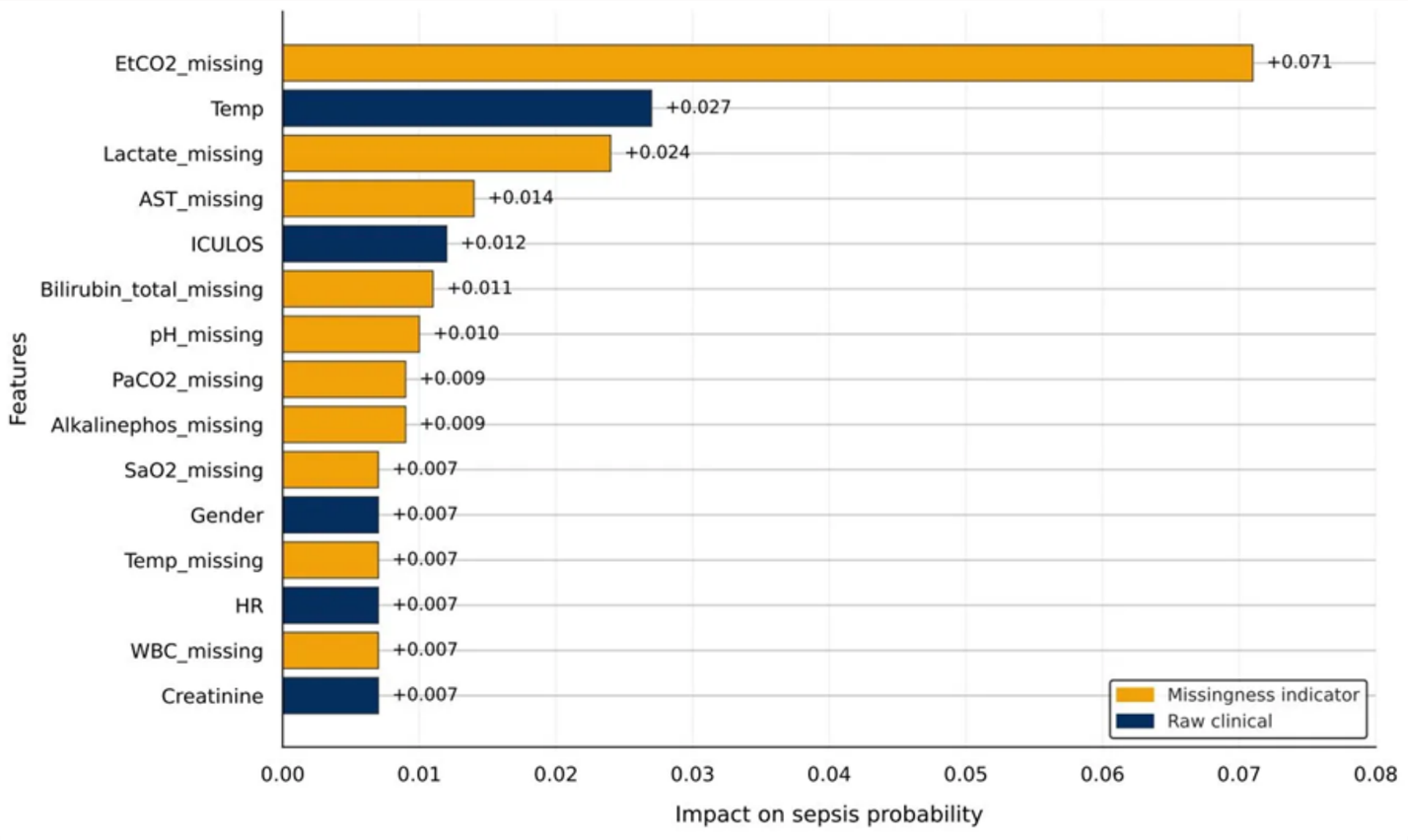


Figure: WinTSR Bar chart – Patient p000161 (Hour 6). Top-15 features

EtCO2_missing dominates by a large margin.

= EtCO2 is collected only in mechanically ventilated patients.

Streamlit Demo

Figure: "Early sepsis prediction" (deployed demo)

Streamlit Demo (2)

Figure: “Early sepsis prediction” (deployed demo)

Conclusions

- **LightGBM with Optuna and gain-based feature selection achieves the best Utility Score (0.381) => Gradient boosting outperforms with structured feature engineering.**
- **Hybrid architecture underperforms standalone LightGBM => General temporal embeddings do not substitute domain-specific aggregates (information bottleneck confirmed)**
- **SHAP and WinTSR independently identify clinically meaningful signals => Cross-architecture consistency confirms models capture real physiological patterns**
- **Additionally, TCN: Competitive results with 2× fewer features, no manual feature engineering (aggregation windows) => Scalability advantage for new clinical settings**

Early prediction of sepsis from time-series clinical data

Anna Rechkalova

Supervisor: Dmytro Kuzmenko,

Sr. lecturer, Dep. of Multimedia Systems