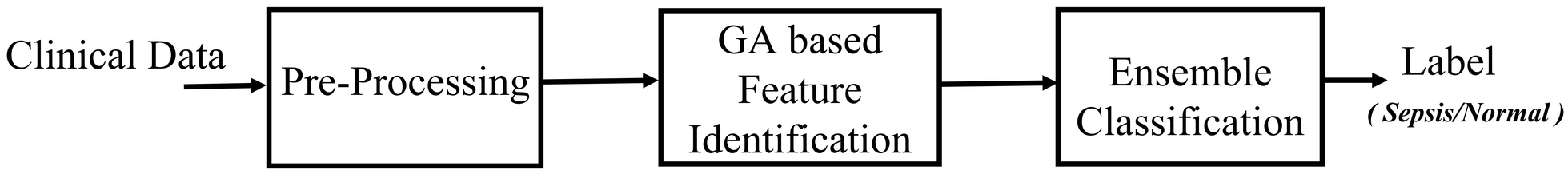
Early Prediction of Sepsis using Ratio and Power-based Features

N Nesaragi and S Patidar
National Institute of Technology Goa, Goa, India
shivnarayan.patidar@nitgoa.ac.in

Paper ID:207 Early Prediction of Sepsis from Clinical Data: PhysioNet/Computing in Cardiology Challenge 2019

Proposed Solution



Influential Features

S.No	Component of Features	Type
1	End Tidal CO_2 / Partial Thromboplastin time (s)	x/y^2
2	Diastolic BP/Gender	x^4/y^8
3	Diastolic BP/Gender	x/y^2
4	Heart Rate/ Age	x/y^2
5	Age / Gender	x/y^2
6	Heart Rate / (Systolic BP * Age)	x/yz
7	Heart Rate (beats per minute)	x^5
8	Temperature (deg C)	x^4
9	Mean Arterial Pressure (mm Hg)	$1/y^2$
10	Diastolic BP (mm Hg)	x^2
11	End Tidal CO_2 (mm Hg)	x^4
12	FiO_2: Fraction of Inspired Oxygen (%)	x^8
13	Alkalinephos (IU/L)	x^6
14	Creatinine (mg/dL)	$1/y^3$
15	Fibrinogen (mg/dL)	x^4
16	Age	x^7
17	ICULOS:ICU Length of Stay	$1/y$

Performance Statistics

Threshold Hours	AUROC	AUPRC	F1-Score	Accuracy	Utility Score
6	85.22	13.32	13.38	87.92	40.00
12	84.00	16.87	18.04	87.73	35.10
18	81.31	11.23	16.21	87.73	30.10
24	80.01	10.12	14.79	87.56	28.51

Performance Statistics of Sepsis Cases

Threshold Hours	AUROC	AUPRC	F1-Score	Accuracy	Utility Score
6	56.13	22.31	28.15	45.01	60.10

Methodology

- **Optimal Sepsis Detection Framework:** Uses GA based optimization and RusBoost classification.
- **Pre-processing:** NaN values are replaced by Linear interpolation.
- **Optimization Objective:** Maximization of Utility Score.
- **Input to Objective Function:** Indices of the possible combinations of the ratio based features only and not the features.
- **Optimization Results:** Identified 17 most influential ratio and power based clinical features.
- **Feature Set:** Given 38 signs plus identified 17 forms the final feature set of 55 values.

Ratio and Power based Features

$$R = \left\{ \frac{x^k}{y^m z^n} : x, y, z \in P; -9 \leq k, m, n \leq 9 \right\}$$

GA based Optimization

$$\theta^* = arg \max_{\theta \in R} \{U_n\}$$

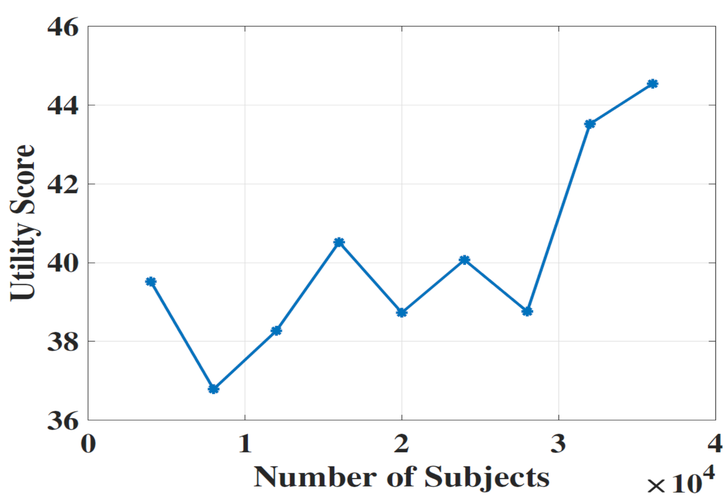
Highlights

- Strength of ratio and power-based features for automated diagnosis of sepsis is explored.
- Method is robust to missing data even at the density of 20% of the data.

Future Work

- To explore feature space further.
- To improve model by adding more training instances.
- To apply faster anomaly detection with identified features.

Effect of Training Data Size



Utility Score

(Test Data Subset)

39.00 %

Run Time

(Test Subset)

15:57:22 in Hours