10. Bootstrap Random Forest Model:

We use the Bootstrap Random Forest model, which is more advanced than the traditional Random Forest algorithm. It operates by training different decision trees on bootstrapped samples (randomly selecting with replacement) of the training dataset. Each tree makes an independent prediction and the predictions are aggregated to form a final prediction.

Key components of the model are:

Random Forest: An ensemble method that takes a large number of trees and decreases the variance that accompanies a singular decision tree.

Bootstrap Sampling: Forming a large number of subsets of the data to train individual trees and therefore increase model stability while also reducing overfitting.

Ensemble Learning: The model offers improved generalization by combining results from all constituent trees, aiding in overcoming individual tree biases.

Classification Task: The model determines if a patient is classified as Septic or Non-septic relying on the clinical features available.

1.Model Architecture

Parameter	Description
Number of Trees	100 decision trees (as per the paper's recommendation).
Maximum Depth	None (trees are allowed to grow without constraints).
Minimum Samples Required for Split	10 (as per the hyperparameters from the paper).
Minimum Samples Required for Leaf	5 (as per the hyperparameters from the paper).
Bootstrap Sampling	Enabled to ensure diversity and reduce overfitting.

2. Model Training

The **Bootstrap Random Forest** model is trained using the preprocessed dataset that has been feature-selected through multiple techniques, such as:

- Random Forest Feature Importance
- Chi-Square and Information Gain (for filter-based selection)

- Recursive Feature Elimination (RFE)
- Forward Selection

Training Process:

1-Preprocessing

2-Splitting the Data

3-Model Initialization

4-Model Training

5-Performance Metrics: The model's performance is evaluated using key metrics such as accuracy, AUC, and F1-Score.

3. Model Evaluation Metrics

We evaluated the Bootstrap Random Forest model using the following metrics:

Accuracy:

-Definition: The proportion of correct predictions out of the total predictions.

-Value for our model: 89.45%.

-Significance: High accuracy suggests that the model is correctly identifying Septic and Non-Septic cases.

AUC (Area Under the Curve):

-Definition: Measures the ability of the model to distinguish between Septic and Non-Septic cases.
-Value for our model: 0.8950.
***Significance: An AUC closer to 1 indicates a better-performing model. Our model's AUC is comparable to Paper 3's 0.908, showing strong discriminatory power.**

F1-Score:

-Definition: The harmonic mean of precision and recall, providing a balanced view of model performance, especially on imbalanced datasets.
-**Value for Class 0 (Non-Septic): 0.88
-**Value for Class 1 (Septic): 0.64
-Significance: The F1-Score for Class 0 is high, indicating that the model is highly precise when predicting Non-Septic cases. However, the F1-Score for Sepsis (Class 1) could be improved further, as it is relatively lower.

Confusion Matrix:

- The confusion matrix shows the breakdown of the model's correct and incorrect classifications for each class (Non-Septic and Septic).

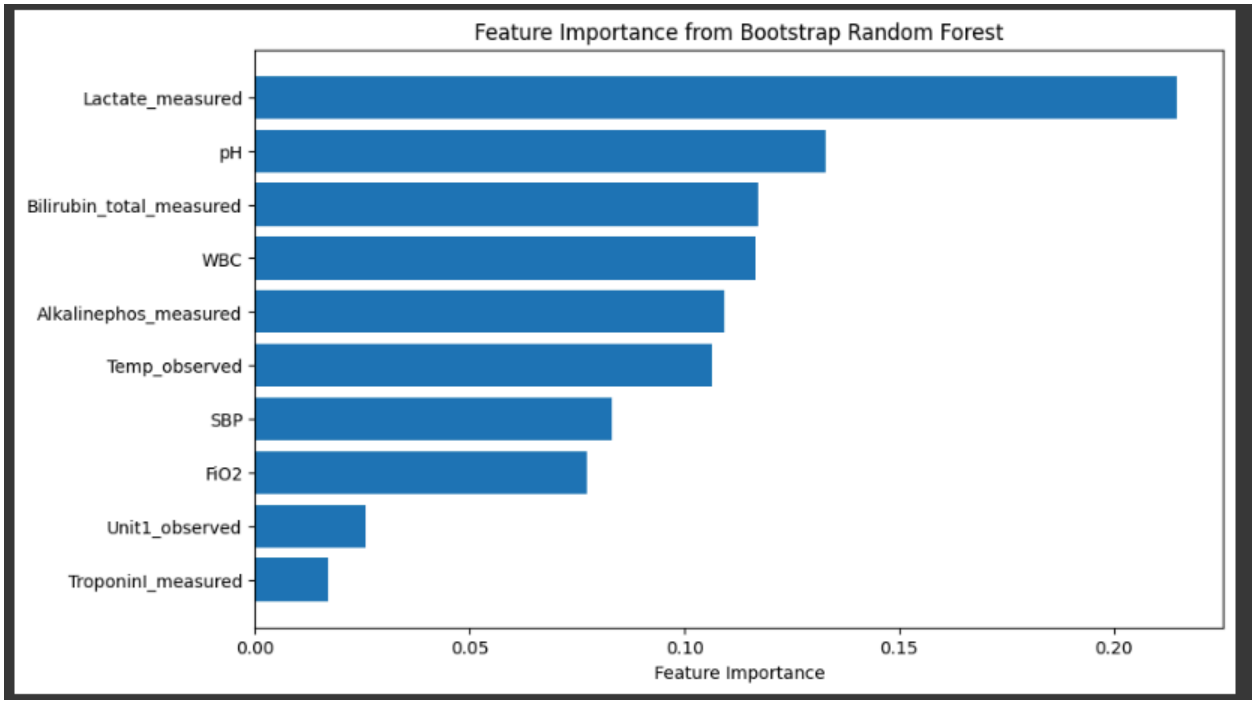
Actual/Predicted	Non-Septic Predicted	Septic Predicted
Actual: Non-Septic	1313 (TN)	87 (FP)
Actual: Septic	274 (FN)	326 (TP)

ROC Curve:

The **ROC Curve** illustrates the trade-off between **True Positive Rate** (Sensitivity) and **False Positive Rate** (1 - Specificity). The **AUC** of **0.8950** indicates that the model is able to differentiate **Septic** and **Non-Septic** cases well.

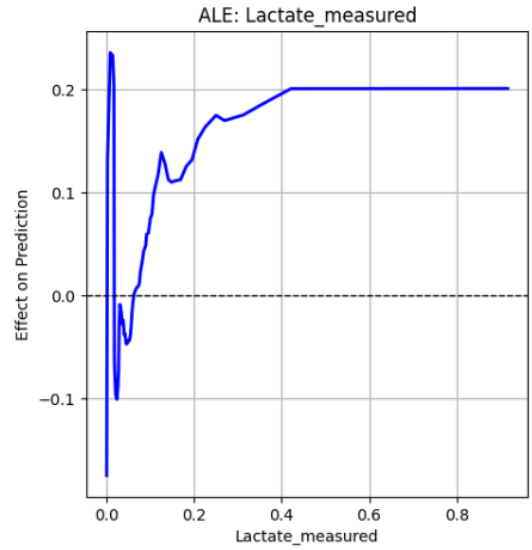
4. Results and Discussion:

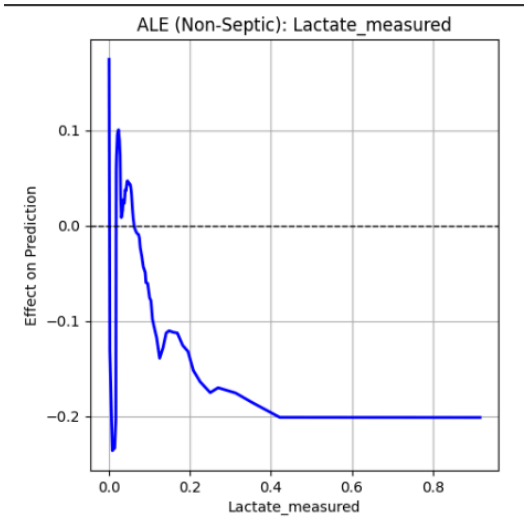
We trained the **Bootstrap Random Forest** model using selected features, identified through feature selection techniques like **Random Forest Importance**, **Information Gain**, **RFE**, and **Forward Selection**. We then used **ALE plots** to assess each feature's impact on predicting **Sepsis** and **Non-Sepsis**. Non-influential features, such as **Unit1 observed** and **AlkalinePhos measured**, were removed, and the model was retrained. The final model, using the remaining important features, demonstrated improved performance with clear interpretability.



These metrics reflect the final performance of the **Bootstrap Random Forest** model after removing non-influential features and retraining with the selected features:

Metric	Value
Accuracy	0.8945
Precision	0.8945
Recall	0.7875
F1-Score	0.8391





In the visualizations, we have two sets of ALE plots: one for **Septic** and one for **Non-Septic** classes. The feature **Lactate_measured** from both sets shows significant effects, especially in **Septic**, where a sharp increase can be seen. For both classes, **Unit1_observed** shows minimal effect, indicating it doesn't significantly influence predictions. By analyzing these curves, we determined which features impact the model's predictions, and the non-influential features were removed before retraining the model. The **Lactate_measured** feature is key in distinguishing Sepsis from Non-Sepsis, and this was confirmed through the ALE visualizations.

Comparison with **Behrad Barghi and Nasibeh Azadeh-Fard:**

Our Model Performance:

From our results using the Random Forest model (with Bootstrap sampling), we achieved the following metrics:

Metric	Our Model (Random Forest)	Paper 3 (Bootstrap Forest)
Accuracy	89.45%	93.5%
AUC	0.8950	0.908
F1-Score	0.64	0.425
Misclassification Rate	0.1055	0.064

Logistic regression

This paper presents a logistic regression-based approach for early detection of sepsis in ICU patients. data from 40,336 ICU patients (PhysioNet Computing in Cardiology Challenge 2019).

Derived Clinical Features

- **Cardiac - Pulse pressure, shock index, cardiac output.**
- **Respiratory - Virtual shunt, Carrico index.**
- **Laboratory - Urea/creatinine ratio, anion gap.**

And apply Z-scores as **Transformations then add it as a new features**

Compersion

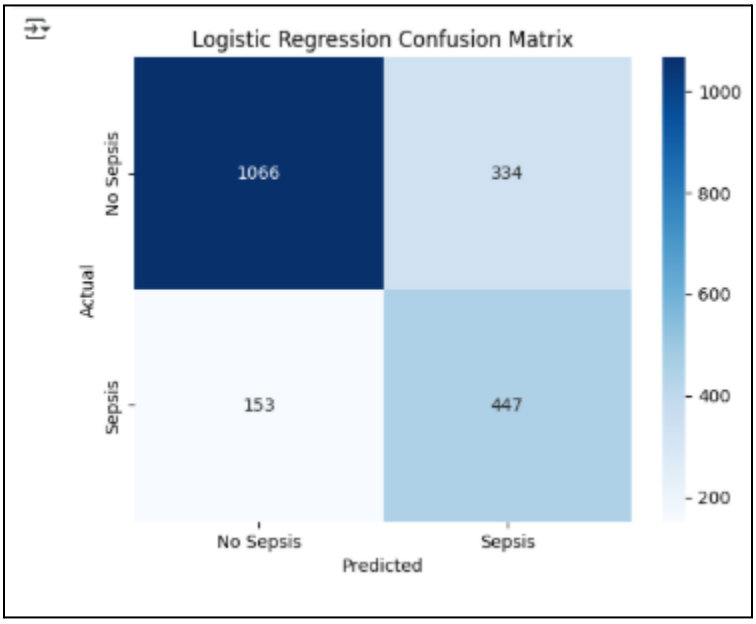
LR With **[Feature Engineering and Z-Score Transformation (implemented in paper)]**

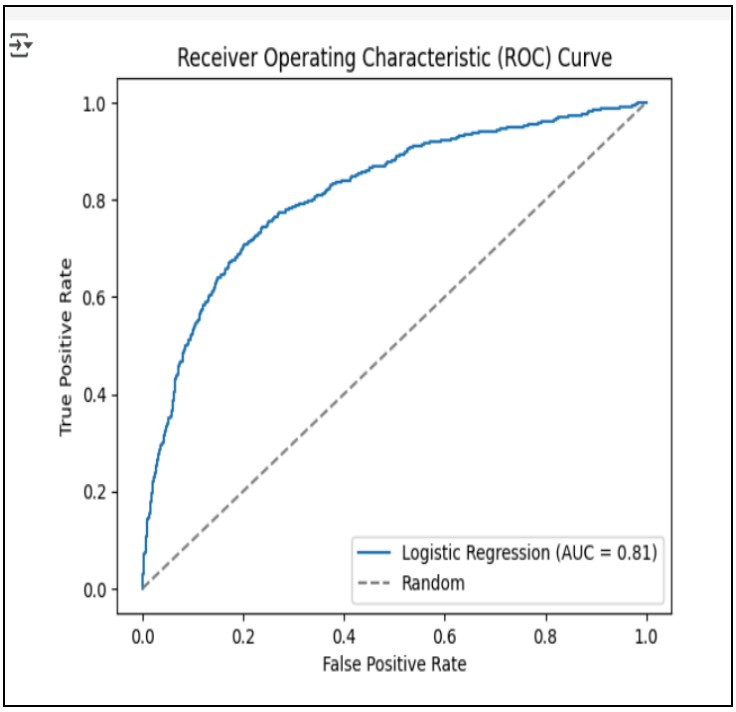
Our findings

- **Precision** for class 0 (non-septic): 0.87
- **Recall** for class 0: 0.76
- **Precision** for class 1 (septic): 0.57
- **Recall** for class 1: 0.74
- **Accuracy**: 0.76

Paper's Findings

- **AUROC** for three test sets: 0.747, 0.760, and 0.783
- **Accuracy** for three test sets: 0.795, 0.889, 0.815





The model shows comparable performance in terms of accuracy (0.76), but there's room for improvement, especially in terms of precision for class 1 and AUROC.

Interpretability Techniques

Feature selected

Selected significant features ['FiO2_observed', 'Lactate_measured', 'PaCO2_observed', 'Phosphate_measured', 'Calcium_observed', 'pH_observed', 'Magnesium_observed', 'Potassium_observed', 'Alkalinephos_measured', 'AST_measured', 'Bilirubin_total_measured', 'PTT_measured', 'BaseExcess_observed', 'BUN_observed', 'Creatinine_observed', 'EtCO2_measured', 'WBC_observed', 'Platelets_observed']

FROM linearity of independent variables with respect to the log odds

Transformation potential: Features with clear non-linear patterns might benefit from transformations, such as taking logarithms or adding polynomial terms to linearize the relationship.

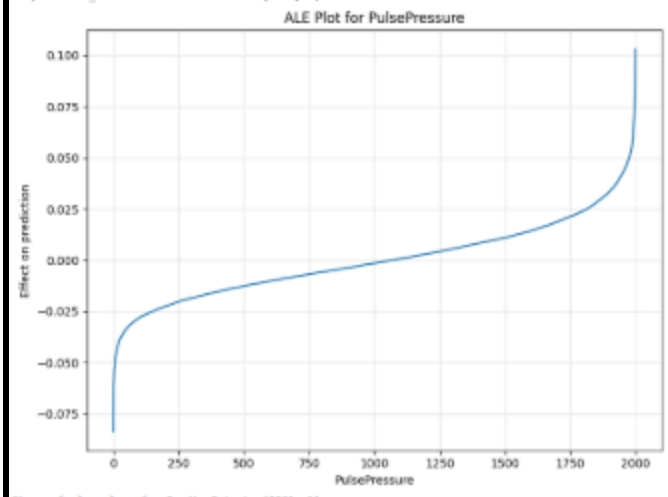
Features like **Blood Pressure**, **Cardiac Output**, and **Shock Index** are not perfectly linear with log-odds and might need further transformations, such as applying a **log-transformation** or **polynomial features** for better model performance.

Based on the Accumulated Local Effects (ALE) plots, here's a summary of how each derived hemodynamic parameter impacts sepsis prediction

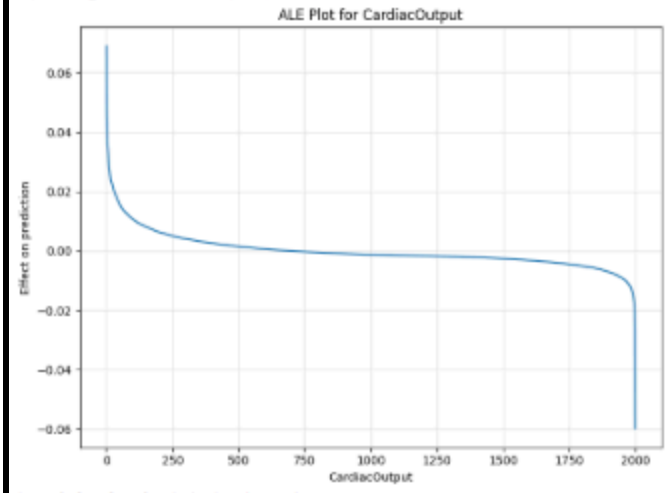
- ➔ **PulsePressure**
  - ◆ Minimal effect at low to moderate values
  - ◆ Positive correlation with sepsis at high values (>1750)
  - ◆ The sharp upward curve at extreme values suggests this is a critical indicator when severely elevated

- ➔ **CardiacOutput**:
  - ◆ Strongest positive effect at very low values
  - ◆ Minimal impact across middle ranges
  - ◆ Negative effect at extremely high values
  - ◆ Suggests both cardiac insufficiency and excessive output may indicate sepsis risk
- ➔ **ShockIndex**:
  - ◆ Nearly flat effect throughout normal ranges
  - ◆ Dramatic positive effect at extreme values
  - ◆ Indicates this parameter becomes highly predictive only in severe cases
- ➔ **ModifiedShockIndex**:
  - ◆ Similar pattern to ShockIndex
  - ◆ Minimal effect across most values
  - ◆ Sharp increase in effect at very high values
  - ◆ Suggests this is primarily useful as a late indicator of sepsis

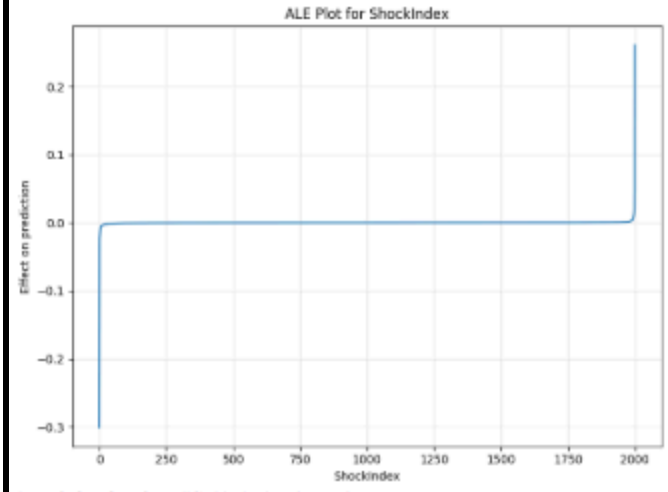
Shape of ale_values for PulsePressure: (2001, 1)



Shape of ale_values for CardiacOutput: (2001, 1)



Shape of ale_values for ShockIndex: (2001, 1)



Shape of ale_values for ModifiedShockIndex: (2001, 1)