Wrightington, Wigan and Leigh NHS Foundation Trust

Lancaster University

Predicting ICU Transfers: A Machine Learning Approach to Early Identification of Critical Patients

Akash Marar

Introduction and Motivation

Brief overview of the problem:

- 30% of mortality at WWL NHS Trust linked to missed diagnoses of rapidly deteriorating patients
- Need for early and accurate prediction of ICU transfers

Importance of predicting ICU transfers:

- Reduce care requirements and associated costs
- Improve patient outcomes
- Enhance hospital resource management

Current limitations of NEWS2 score:

- Does not capture complex interactions between patient variables
- Misses some patients who deteriorate rapidly

Key Objectives

- Develop a predictive model to accurately forecast ICU transfers using machine learning techniques.

- Identify the most important features that contribute to ICU transfer predictions for better clinical decision-making.

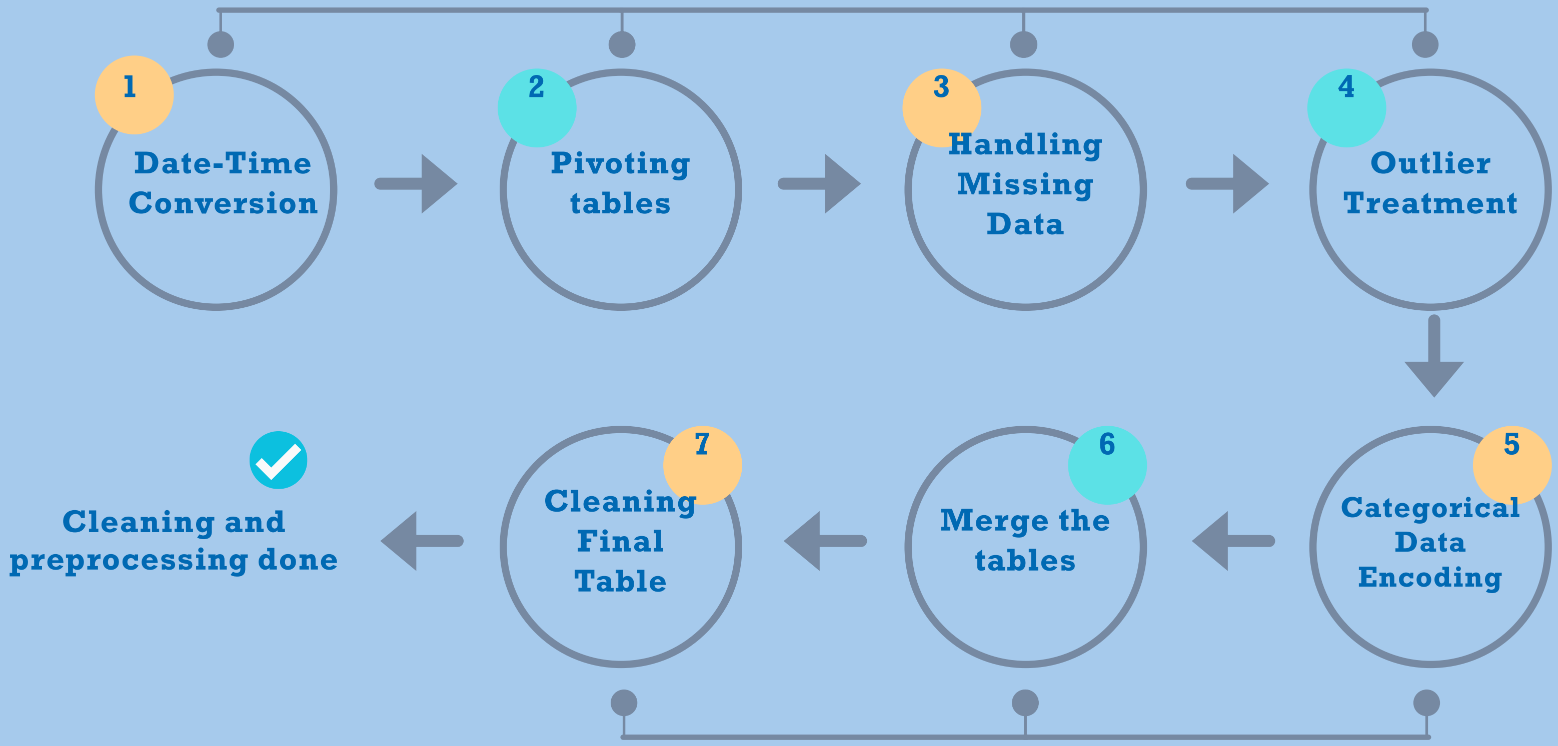
Dataset Overview

1. Target Dataset • Patient admission and outcome data
2. Observe Dataset • Patient vital signs and NEWS2 components
3. Diagnostic Dataset • Results of various diagnostic tests

Data Characteristics:

- Time span: Jan 2019 - May 2024
- 73,628 unique patients all aged 18+

Simplified steps for Preprocessing



Date-Time Conversion and Pivoting tables

1.Date-Time Conversion • Converted AdmitDtm and Obs_dtm to datetime format • Used pandas' to_datetime() function • Applied 'errors='coerce'' to handle inconsistencies • Enabled accurate time-based calculations and filtering

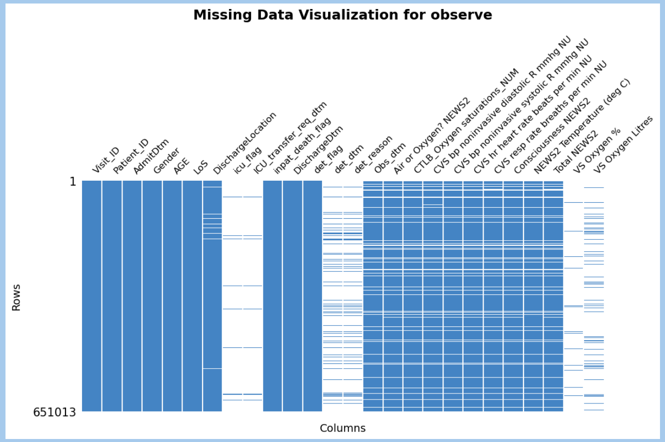
2.Pivoting • Transformed data into wide-format table • Each row: unique combination of Patient ID, Visit ID, AdmitDtm, Obs_dtm • Results became columns

Key Benefits:

•Improved data consistency

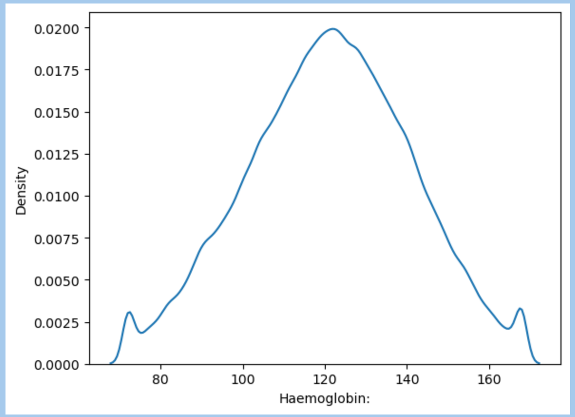
•Enhanced time-based analysis capabilities

•Optimized format for a model as each unique test will become a feature



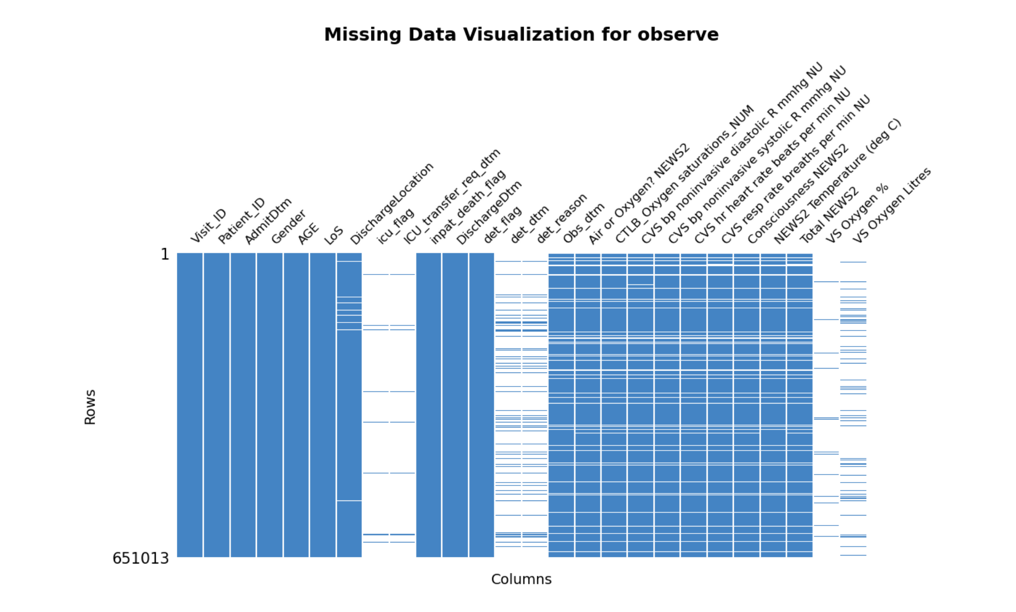
Outlier Treatment and Categorical Data Encoding

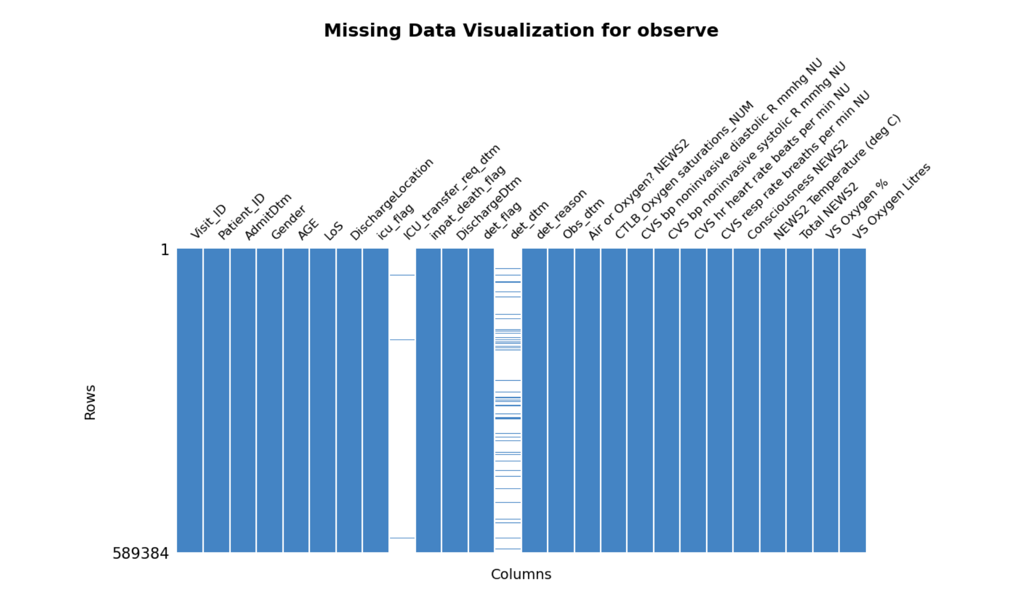
1. Outlier Treatment: Winsorization • Applied to numerical columns • Replaced values below 1st and above 99th percentiles • Balanced approach: preserves data structure while reducing outlier impact • Suitable for medical data: retains potentially significant extreme values Benefits: • Maintains data distribution • Preserves clinically relevant information • Improves model resilience to extreme values
2. Categorical Data Encoding • Transformed categorical data into numerical format Importance: • Enables inclusion of categorical variables in models • Preserves categorical information in a model-compatible format



Handling Missing Data

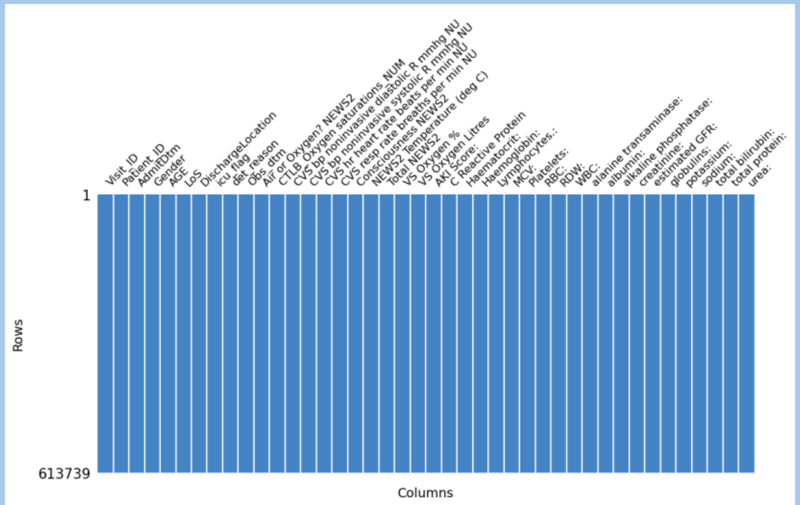
- Data Filtering: Removed rows missing critical timestamp data .
- Imputation: Applied forward and backward filling for time-series data, and iterative imputation to estimate remaining missing values.
- Duplicate and Threshold Filtering: Removed duplicate rows and filtered rows with more than 16 missing values.
- Data Validation: Validated binary variables, corrected outliers, capped percentages, and ensured all values were clinically possible.





Data Combination and Final Preprocessing

1. Data Integration • Merged observation and diagnostic datasets • Aligned datetime fields across datasets • Performed outer join for comprehensive patient view
2. Handling Missing Data • Conducted thorough missing value analysis • Applied forward/backward filling for time-series data • Used IterativeImputer for remaining missing values
3. Data Cleaning • Removed rows with missing crucial timestamps • Filtered out rows with excessive missing values • Eliminated duplicate entries
4. Preventing Data Leakage • Removed post-event variables (e.g., DischargeDtm, ICU_transfer_req_dtm)
5. Data Validation and Correction • Ensured correct data types and value ranges • Rounded scores and counts to whole numbers • Removed impossible values (e.g., negative bilirubin)



Feature Selection and Data Preparation

Feature Selection

- Included: • Patient demographics • Vital signs • Diagnostic test results
- Excluded to Prevent Data Leakage: • Identifiers (e.g., Visit ID, Patient ID) • Future information (e.g., DischargeDtm, ICU_transfer_req_dtm) • Redundant information (e.g., AdmitDtm)

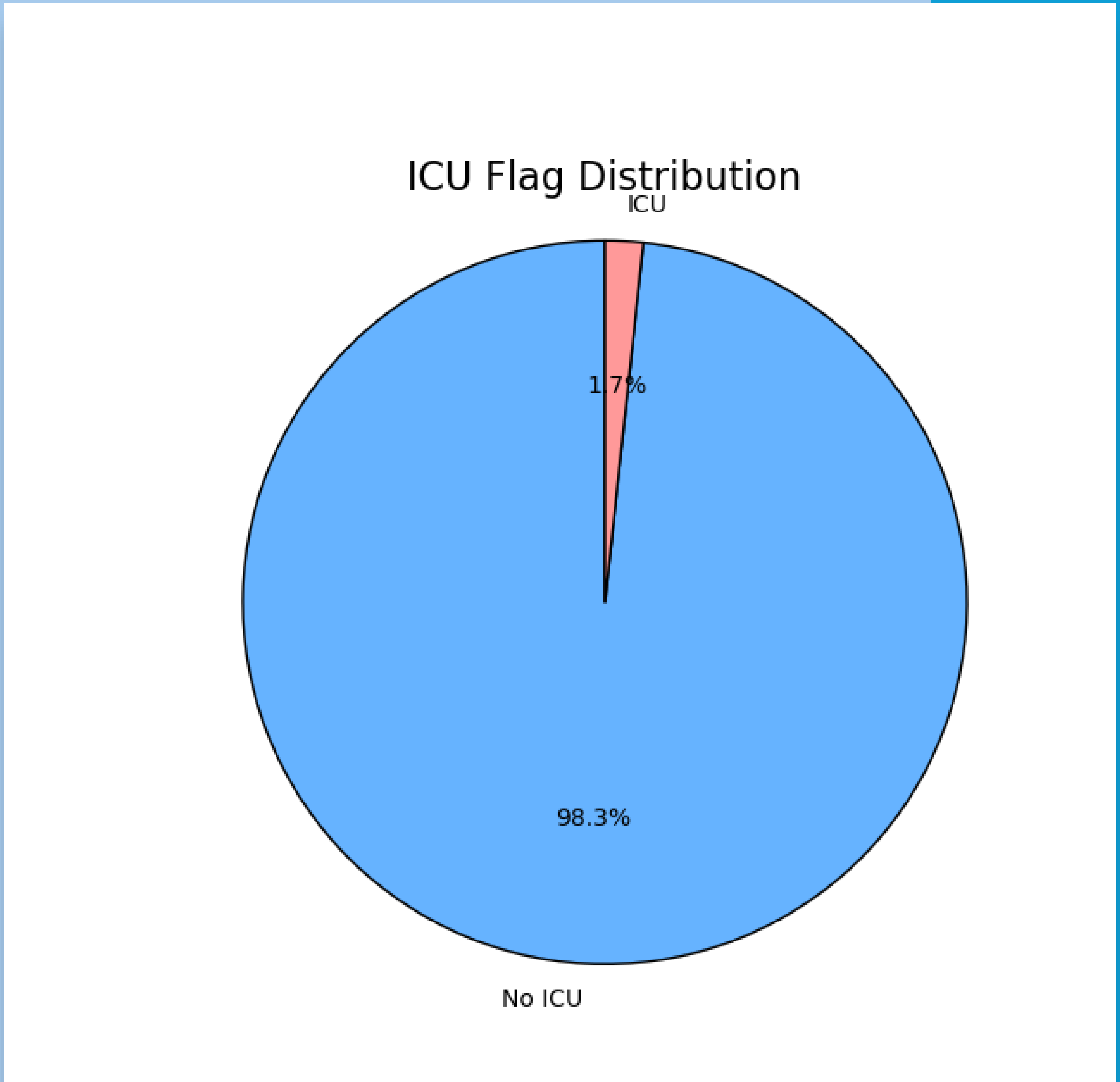
1. Target Variable • 'icu_flag': Binary indicator of ICU transfer
2. Data Type Consistency • Ensured appropriate data types for all features • Encoded categorical variable gender as mentioned before

Key Principles:

- Exclusion of post-event information
- Focus on real-time available data
- Prevention of overfitting and data leakage to maintain the integrity of predictive task

Class Imbalance

- Only 1.7% of patients required ICU transfer, highlighting a major class imbalance.
- Imbalance makes models biased toward predicting "No ICU" (98.3% majority).
- To improve predictions, various balancing techniques need to be tested



Addressing Class Imbalance

Techniques Explored:

1. Oversampling Methods • **SMOTE**: Creates synthetic minority examples • **ADASYN**: Focuses on samples near decision boundary • **Random Over-Sampling**: Duplicates minority samples
2. Undersampling Methods • **Random Under-Sampling**: Removes majority samples • **Tomek Links**: Removes noisy majority samples • **Edited Nearest Neighbors (ENN)**: Cleans both classes • **Near Miss**: Keeps relevant majority samples
3. Combination Method • **SMOTE + Random Under-Sampling**

Rationale for Multiple Techniques:

- Different methods suit various aspects of imbalance
- Allows comparison of effectiveness in ICU transfer prediction

Model Selection

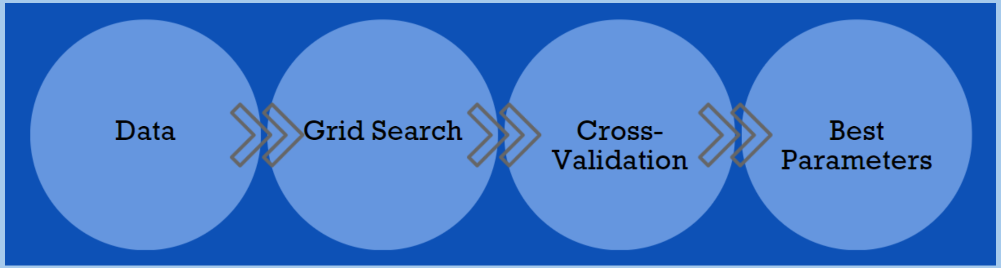
Model Selection: Random Forest and XGBoost

1. Random Forest • Ensemble method Good generalization, reduced overfitting
2. XGBoost • Captures complex, non-linear relationships • High performance crucial for medical predictions
3. Combined Strengths • Leverages two different ensemble approaches • Creates a comprehensive predictive framework • Enhances overall predictive capability for ICU transfers

Benefits of Using Both: • Robust performance across different data characteristics • Complementary strengths for complex medical data • Increased confidence in predictions through model comparison

Hyper Parameter Tuning

1. Importance of Hyperparameter Tuning
    - Optimizes model performance for specific datasets
    - Addresses overfitting:
    - Prevents models from memorizing training data
    - Improves generalization to unseen data



2. Methodology
    1. Utilized sklearn's GridSearchCV which:
    2. Systematically works through multiple combinations of parameter tunes
    3. Trains and evaluates a model for each combination •
    4. Implemented 3-fold cross-validation:
    5. Helps in assessing model performance consistency
    6. Optimized for F1-score

3. Tuned Parameters
    - Random Forest: • n_estimators • min_samples_split • min_samples_leaf • max_features • bootstrap
    - XGBoost: • n_estimators • learning_rate • subsample • colsample_bytree • scale_pos_weight

Temporal Split and Model Validation

1. Data Split • Training and Initial Testing: Data before April 1, 2024 • Validation Set: Data from April 1, 2024 onwards
2. Model Validation • All models evaluated on post-April 2024 validation set • Includes:
    1. Baseline model
    2. Models with different imbalance handling techniques
3. Rationale • Simulates real-world application • Trains on historical data, predicts on future unseen data • Tests model performance in practical scenarios

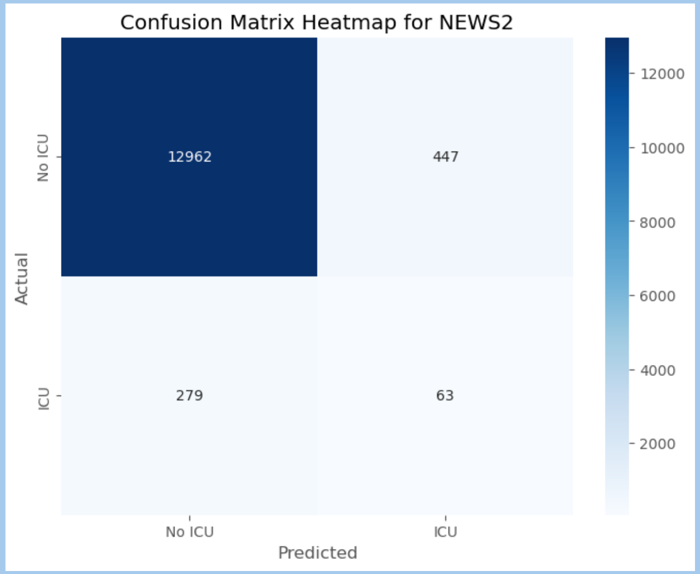
Key Benefits:

- Realistic assessment of model generalization
- Avoids data leakage from future to past
- Evaluates model's ability to handle temporal shifts in data

Performance of Total NEWS2 Score

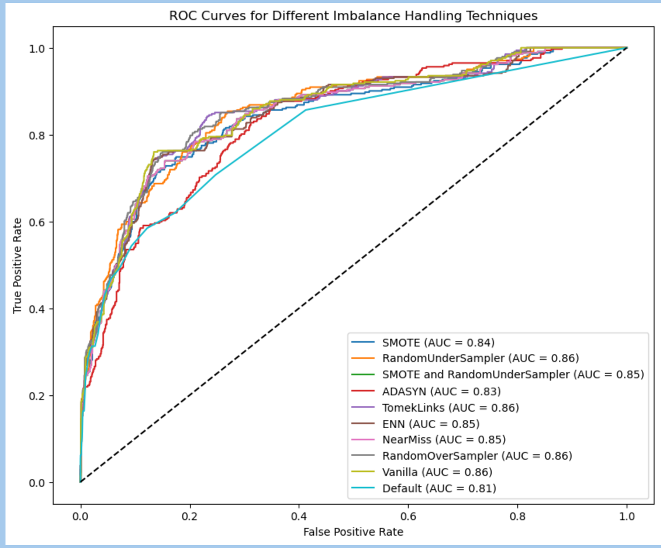
Performance Metrics:
•Precision: 0.12
•Recall: 0.18
•F1-score: 0.15

Interpretation:
•Only 12% of patients flagged for ICU transfer actually required it
•NEWS2 captured only 18% of patients who truly needed ICU transfer
•High number of false positives and false negatives indicates significant room for improvement, which our model aims to address

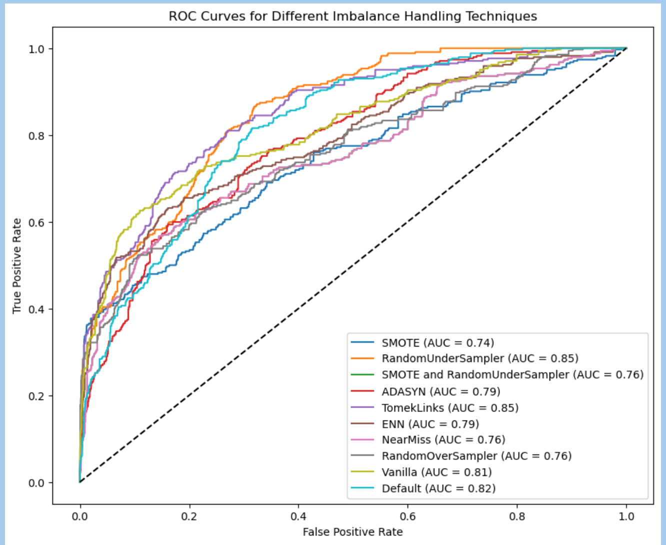


ROC Curve Analysis: Random Forest vs. XGBoost

- **Random Forest** shows more **consistent performance** across different imbalance handling techniques, with AUC values clustering between **0.83 and 0.86**.
- **XGBoost** results are more varied, with AUC values ranging from **0.74 to 0.85**.
- Techniques like **RandomUnderSampler** and **TomekLinks** perform well with both models, but **Random Forest** offers a **more stable ROC** across techniques.



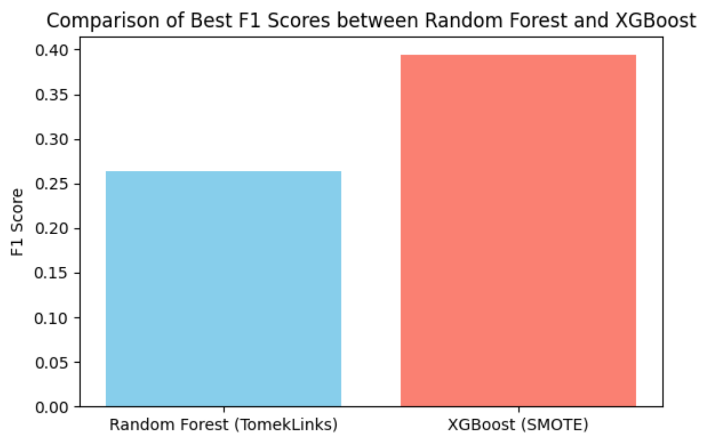
Random Forest



XGBoost

F1-Score and Best Performing Model

- While ROC curves show that Random Forest consistently performed better, we chose the F1-score as our primary metric.
- The F1-score was selected to balance precision (avoiding unnecessary ICU transfers) and recall (capturing patients in need of ICU care)
- The best results were achieved using XGBoost with SMOTE and Random Forest with TomekLinks.
- **XGBoost with SMOTE** delivered the highest F1 score, making it the most effective model



Feature Importance Analysis: Default XGBoost vs SHAP

Key Observations:

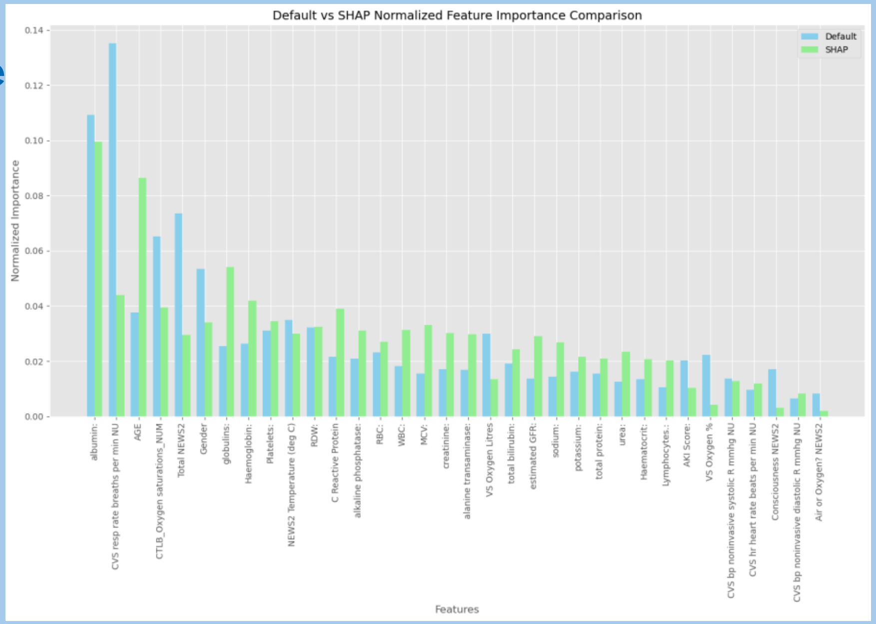
- Both methods agree on high importance of albumin levels
- SHAP assigns higher importance to age (AGE)
- Default method emphasizes oxygen-related features more

Preference for SHAP Method:

- Grounded in cooperative game theory
- Accounts for feature interactions
- Captures complex, non-linear relationships

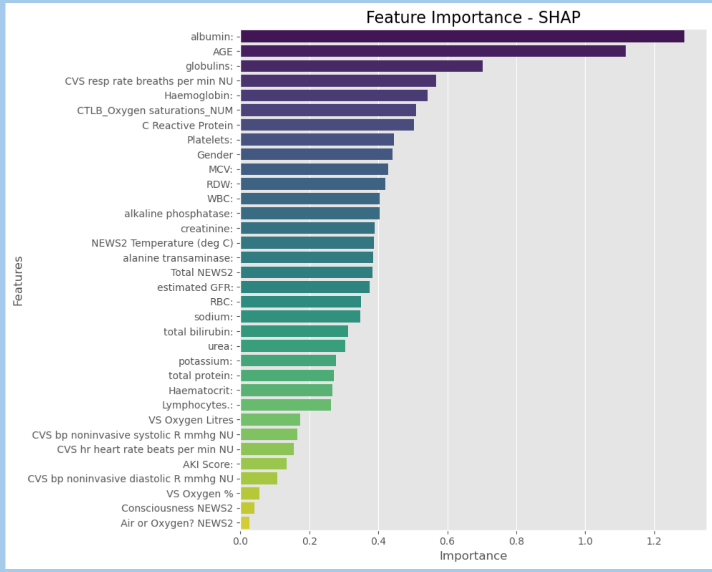
Implications:

- Albumin levels are crucial predictors across methods
- Age may have more complex impact than initially thought
- Oxygen-related features' importance varies by method



Feature Importance Insights

- Key Features: albumin, age and globulin are the top predictors.
- SHAP Analysis: SHAP reveals the non-linear impact of features like age, which ranks higher than in traditional methods.
- Balanced Data: Vital signs and lab values both play crucial roles in ICU transfer predictions.
- Clinical Relevance: Combining easily measurable bedside data with lab results provides a balanced and actionable model for ICU transfer predictions.



Limitations and Ethical Considerations

Limitations:

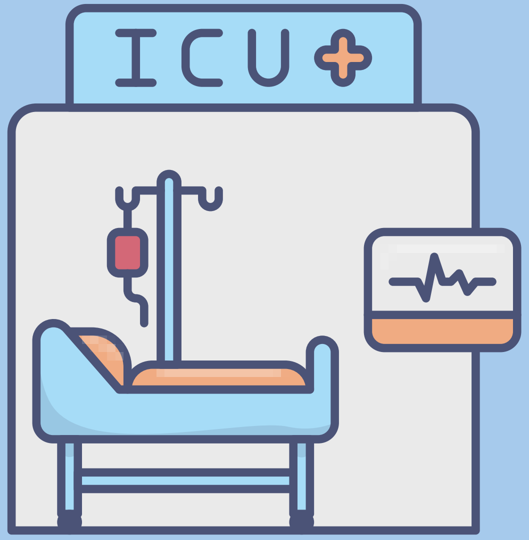
1. Data Constraints • F1 score of 0.39 indicates room for improvement • Lack of doctors' diagnostic notes in dataset • Need for more complex clinical decision-making data
2. Validation Needs • Further testing required on data from other hospitals • Live hospital data testing needed
3. Implementation Challenges • Integration into existing workflows and IT systems

Ethical Considerations:

1. Model Transparency • Accessible documentation of architecture and decision-making
2. Fairness • Regular audits to prevent discriminatory predictions
3. Human Oversight • Model should complement, not replace clinical judgment • Training for healthcare providers on model use and limitations
4. Data Privacy • Adherence to UK data protection regulations • Robust anonymization measures
5. Accountability • Clear framework for model-influenced decisions
6. Patient Communication • Transparent information about model's role in care

Future Work

1. Incorporate Doctors' Diagnostic Notes • Use natural language processing to extract insights • Capture important observations not in structured data
2. Collaborate with Healthcare Professionals • Discuss significant features identified by the model • Refine data collection based on clinical expertise • Recommend frequency of measurements for key indicators
3. Explore Alternative Modeling Approaches • Investigate time-series analysis • Capture progression of patient conditions over time
4. Expand Data Collection • Include a wider range of hospitals • Improve model generalizability • Account for variations in patient populations and resources
5. Enhance Model Interpretability • Develop more intuitive ways to present model predictions • Create user-friendly interfaces for clinical staff



Final Remarks

Key Achievements:

- XGBoost with SMOTE outperformed NEWS2 in ICU transfer prediction
- Addressed class imbalance effectively
- Enhanced model interpretability using SHAP

Potential Impact:

- Improved critical care management
- Optimized resource allocation
- New insights for patient monitoring and risk assessment

This study serves as a stepping stone towards enhancing patient care through ethically implemented, data-driven decision support systems in ICU predictive modeling.

Wrightington, Wigan and Leigh NHS Foundation Trust

Lancaster University

Acknowledgements

•Nishchay Joshi for his invaluable supervision and constant support
•Thomas Ingram and Brian Wood for their insightful contributions and guidance
•Marco Battiston, my university supervisor, for his excellent advice and timely support throughout the project and dissertation
•Wrightington, Wigan and Leigh NHS Foundation Trust for commissioning this project
•Lancaster University for providing the resources and environment to conduct this research