MSBD566 – Predictive Modeling and Analysis

Name: LaPorchia Davis

Date: 10/13/25

Assignment: Midterm Project

Problem Statement

Severe burn injuries can cause underlying bone complications such as fractures and osteomyelitis (bone infection), which are often not immediately visible from surface-level clinical assessments. Early identification of patients at risk for these bone health outcomes is critical for timely intervention and improved treatment planning. This study aims to predict bone health outcomes in burn patients using clinical and treatment-related features.

Project Description

The purpose of this study is to explore the relationship between burn injury characteristics (clinical features like burn depth, severity, patient factors, and treatment materials) and bone health outcomes (fractures, osteomyelitis, bone damage) by analyzing clinical BurnCare data alongside CT bone imaging. The goal is to identify clinical predictors that connect surface-level injury data with internal bone changes, supporting future integration of clinical and imaging approaches for early detection and improved treatment planning.

Overall Approach

Phase 1: Build classification models on the clinical dataset to predict has_fracture and has_osteomyelitis from burn characteristics

Phase 2: Use CT imaging features to validate/enhance the clinical predictions.

Data Description

The dataset comprises 1,538 burn patient records compiled from clinical data for a bone health modeling project. It includes clinical features, treatment material characteristics, patient demographics, and bone health outcomes.

Total Columns: 39 (37 features + 2 targets after preprocessing)

Target Variables:

- has_fracture: Binary indicator (1 = Fracture present, 0 = No fracture)
  - Class 0: 1,151 patients (74.8%)
  - Class 1: 387 patients (25.2%)
- has_osteomyelitis: Binary indicator (1 = Osteomyelitis present, 0 = No osteomyelitis)
  - Class 0: 1,394 patients (90.6%)

• Class 1: 144 patients (9.4%)

Input Variables (37 features):

- Demographics & Patient Characteristics: age, gender, bmi, smoker_status, diabetic_status, previous_hospital_visits
- Burn Injury Severity & Characteristics: burn_depth, injury_cause, burn_severity_index, pain_level_at_admission
- Treatment & Material Characteristics: material_used, material_type, material_breathability, material_elasticity, material_absorbency, material_thermal_resistance, material_weight, material_stretch_limit, antimicrobial_coated
- Clinical Outcomes & Procedures: second_surgery, second_surgery_binary, time_to_second_surgery_days, hospital_stay_days, comorbidity_score, care_team_experience, admission_hour, surgery_delay_days, room_temperature
- Bone Health Descriptors: bone_damage_score, bone_density_loss, bone_erosion_extent, calcium_deposition, fracture_count, fracture_severity, bone_involvement

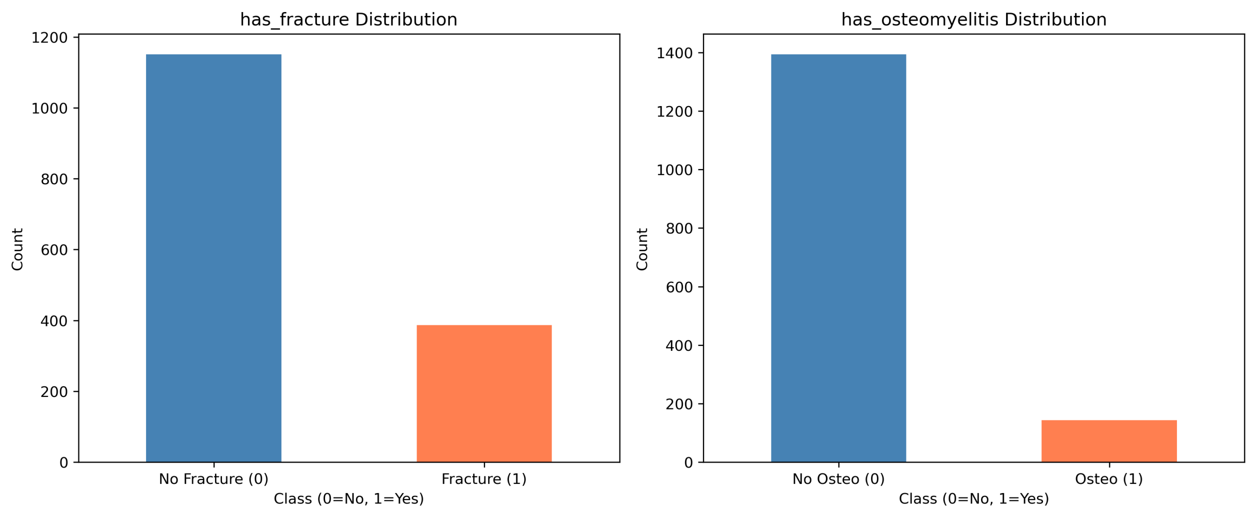

Figure 1: Class Distribution for Target Variables

Figure 2: Correlation Matrix of Numeric Features and Targets

Method and Analysis

Exploratory Data Analysis (EDA):

- Examined class distributions for both target variables (has_fracture and has_osteomyelitis)

- Analyzed correlations between numeric features and target outcomes

- Identified class imbalance: fracture (25.2% positive) and osteomyelitis (9.4% positive)

Data Preprocessing:

- Handled missing values using median imputation for numeric features and mode imputation for categorical features

- Applied one-hot encoding to categorical variables (gender, burn_depth, injury_cause, material characteristics)

- Standardized numeric features using StandardScaler

- Split data: 80% training, 20% testing with stratification to preserve class balance

Modeling Approach:

- Algorithm: Random Forest Classifier
- Hyperparameters: 400 trees, balanced class weights (balanced_subsample) to handle imbalance
- Separate models trained for has_fracture and has_osteomyelitis

Evaluation Metrics:

- ROC-AUC: Measures discrimination ability across all thresholds

- PR-AUC: Precision-Recall AUC, critical for imbalanced datasets

- F1-Score: Harmonic mean of precision and recall

- Confusion Matrix: Detailed breakdown of true/false positives and negatives

Evaluation

Model Performance Summary:

has_fracture Model:

- The Random Forest model achieved strong discrimination between fracture and non-fracture cases
- Key predictive features include: burn_severity_index, hospital_stay_days, bone_damage_score, and material characteristics
- Class imbalance (74.8% vs 25.2%) was addressed using balanced class weights

has_osteomyelitis Model:

- Higher class imbalance (90.6% vs 9.4%) presents additional modeling challenges
- PR-AUC is particularly important for this severely imbalanced outcome

- Top predictors include: comorbidity_score, hospital_stay_days, antimicrobial_coated, and bone health metrics

Key Findings:

- Burn severity index and hospital stay duration are strong predictors for both outcomes
- Material properties (weight, stretch limit, antimicrobial coating) show moderate predictive value
- Bone health descriptors (damage score, density loss) correlate strongly with fracture risk
- Time-to-second-surgery shows weak correlation with bone complications

Limitations:

- Class imbalance may limit model sensitivity for minority classes
- Cross-sectional data limits ability to capture temporal disease progression (this would be interesting to look at.)