MSBD566 – Final Project Report

Predictive Modeling and Analysis
Name: LaPorchia Davis
Date: December 7th, 2025

Integrating Clinical Predictive Modeling and Imaging-Based Analysis to Identify Bone Health Complications in Burn Patients

Project Description

Severe burn injuries often affect more than the surface layers of the skin. In many cases, patients develop **underlying bone complications** including fractures, bone density loss, and osteomyelitis that cannot be immediately detected through visual inspection or standard burn assessments. Early identification of patients at high risk for these complications is crucial, as delayed recognition can negatively impact treatment strategies, surgical planning, and long-term recovery outcomes.

This final project builds upon the earlier midterm work by expanding the analytical pipeline into a **multimodal framework** that integrates:

1. **Clinical predictive modeling**,
2. **Dimensionality reduction using PCA**, and
3. **Deep-learning–based UMAP analysis of CT imaging**.

Together, these stages aim to uncover how both clinical and imaging-derived features contribute to underlying bone health outcomes in burn patients. The central goal is to understand whether surface-level injury characteristics, treatment decisions, and embedded imaging patterns can collectively support early detection of fractures and osteomyelitis.

Data Description

The dataset used for this project consists of **1,538 burn patient records**, each containing a rich mix of demographic information, burn characteristics, treatment properties, clinical outcomes, and bone health descriptors MSBD566_Assignment3_Report (2).

There are **37 predictor variables** and **2 primary target variables**:

- **has_fracture** (Binary: 1 = fracture present)
    - Class balance: 25.2% positive, 74.8% negative
- **has_osteomyelitis** (Binary: 1 = osteomyelitis present)
    - Class balance: 9.4% positive, 90.6% negative

Predictor variables encompass:

- **Demographics:** age, gender, BMI, comorbidity score, smoking and diabetic status
- **Burn injury descriptors:** burn severity index, burn depth, injury cause, pain level
- **Treatment and material characteristics:** antimicrobial-coated materials, breathability, elasticity, thermal resistance, stretch limit, weight
- **Procedural timing:** hospital stay days, time to second surgery, second surgery status
- **Bone health descriptors:** bone damage score, density loss, erosion extent, fracture severity, fracture count

A correlation matrix included in the earlier midterm analysis (page 3) highlighted several strong associations, including the links between **burn severity**, **hospital stay duration**, and **bone health metrics**, all of which guided the modeling strategy MSBD566_Assignment3_Report (2).

Method and Analysis

Preprocessing and EDA

The analysis began with exploratory data evaluation to understand class distributions, numerical ranges, and relationships within variables. Because the osteomyelitis class was severely imbalanced (only 9.4% positive), additional care was taken in model selection and performance evaluation.

Missing values were handled systematically: numeric variables were imputed using the median, and categorical variables using the mode. Categorical predictors were transformed using **one-hot encoding** to prepare them for machine learning algorithms. All numeric variables were standardized using **StandardScaler**, which was essential both for PCA and for ensuring equal weighting among features. A stratified 80/20 train-test split preserved class proportions across subsets.

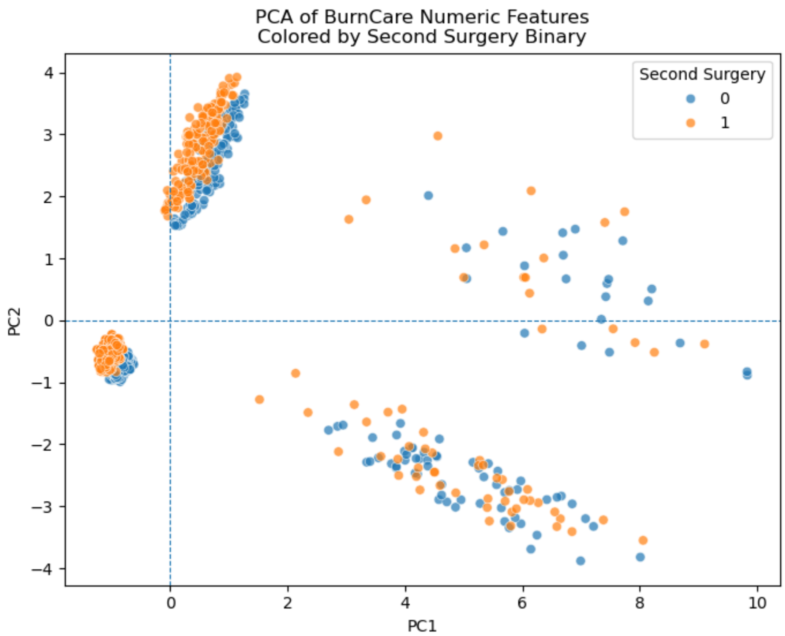
Clinical Predictive Modeling

A **Random Forest Classifier** was selected for both prediction tasks (fracture and osteomyelitis) due to its interpretability, robustness, and ability to model nonlinear relationships. The model used **400 trees** and balanced class weights (`balanced_subsample`) to address outcome imbalance.

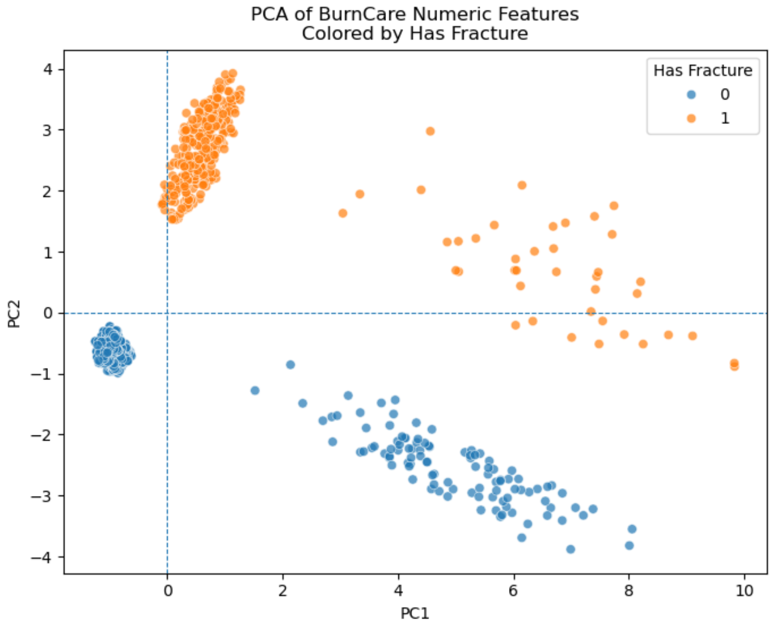
Evaluation used a combination of ROC-AUC, PR-AUC, F1-score, and confusion matrices. The **fracture model** performed well, identifying burn severity index, bone damage metrics, and hospital stay as the strongest contributors. Treatment material properties particularly weight and stretch limit also played an important role, suggesting biomechanical interactions between dressings and underlying bone conditions. For **osteomyelitis**, model performance was more modest due to extreme class imbalance, but predictors such as comorbidity score, antimicrobial-coated materials, bone density loss, and hospital stay emerged as meaningful indicators.

Principal Component Analysis (PCA)

PCA was used to extract low-dimensional structure from the clinical dataset. The analysis showed that **PC1 corresponded to a "severity–bone damage" axis**, with high loadings from burn severity index, bone damage score, fracture count, and length of hospitalization. **PC2 represented a "treatment–comorbidity" axis**, influenced by comorbidity score, second surgery, and antimicrobial-coated dressings.



Plotting PC1 against PC2 revealed that patients with fractures clustered more tightly along the higher PC1 region, reflecting underlying structural involvement. Osteomyelitis cases appeared more dispersed, consistent with their rarity. PCA thus provided an interpretable framework connecting clinical severity with deeper bone outcomes.



UMAP Imaging Analysis Using CNN-Based Feature Extraction

In addition to clinical modeling, the project incorporated an unsupervised imaging analysis to examine whether CT slice-level bone patterns correspond to meaningful anatomical or pathological differences.

Introduction

Dimensionality reduction is a powerful tool in medical imaging because it enables the visualization of high-dimensional data in a low-dimensional space. In this project, a combination of **ResNet-50 CNN feature extraction** and **UMAP projection** was used to analyze 72 DICOM slices from a right knee CT scan. The purpose of this analysis was not to diagnose injuries but to explore whether CNN-based representations reveal distinct anatomical groupings within the scan volume. This helps demonstrate how unsupervised learning can uncover structure that may later support injury detection or clinical decision-making.

Methods

DICOM Preprocessing

All 72 slices were loaded using *pydicom*, applying the VOI LUT when available to preserve radiological contrast. Each image was normalized to 0–1 intensity, converted to grayscale, then expanded to three channels to meet ResNet-50 requirements. Images were resized to **224 × 224**, normalized using ImageNet parameters, and prepared for feature extraction.

CNN Feature Extraction

A pretrained ResNet-50 network was used strictly as a **feature extractor**, not for classification. The final classification layer was removed, and the **global average pooling layer output (2048-dimensional vector)** served as the slice's embedding. These vectors capture bone texture, soft tissue density, cartilage shape, trabecular architecture, and cortical morphology.

UMAP Dimensionality Reduction

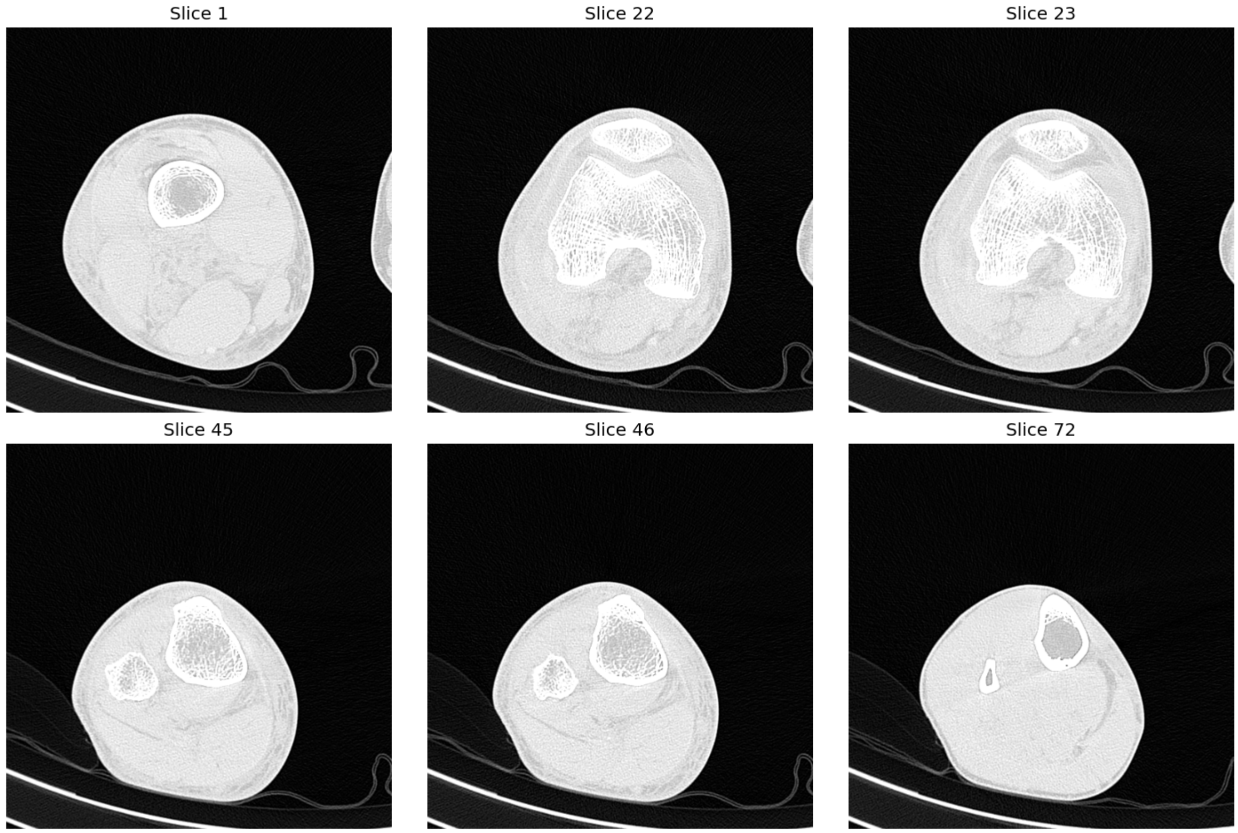
UMAP was applied to the resulting 72 × 2048 matrix using:

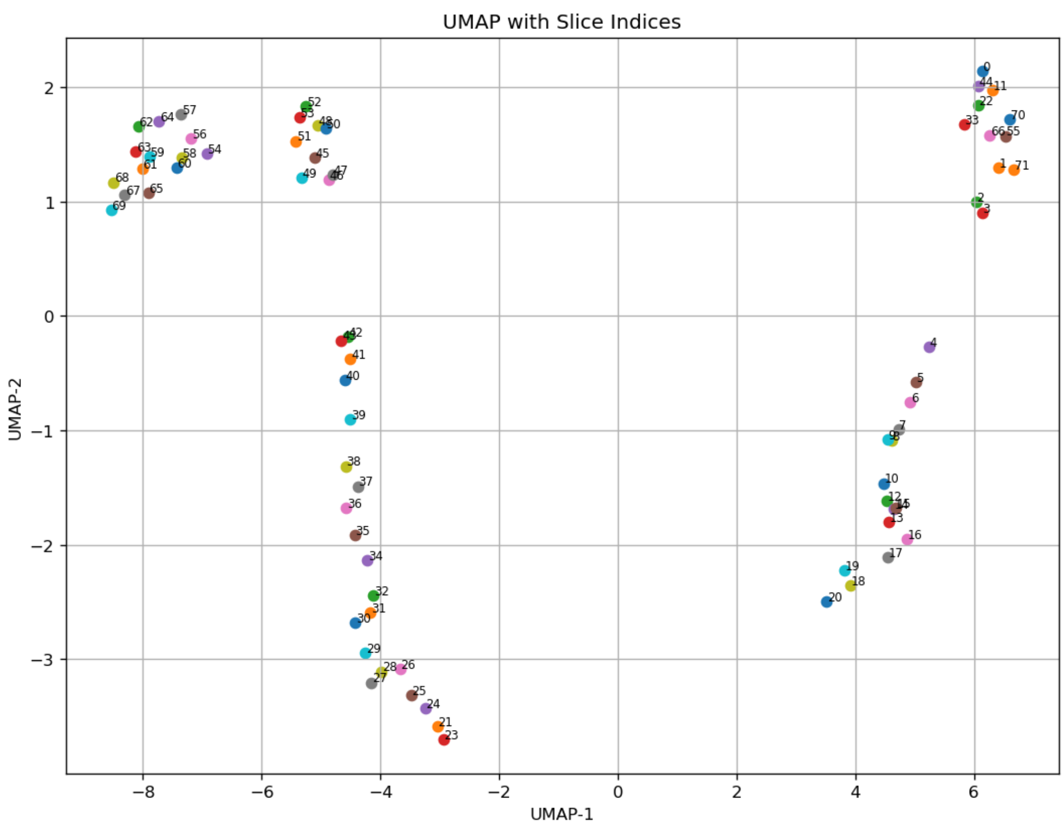
- Metric: Euclidean
- n_neighbors: 10
- min_dist: 0.1
- random_state: 42

These parameters emphasize preservation of local slice-to-slice structure. The 2D embedding was visualized with slice indices annotated.

Slice Interpretation

Representative slices 1, 22, 23, 45, 46, and 72 were examined to interpret anatomical differences and cluster boundaries across the embedding.


Slice 1
Slice 22
Slice 23
Slice 45
Slice 46
Slice 72


UMAP with Slice Indices
UMAP-1
UMAP-2

Results

UMAP Clusters Reflect Anatomical Regions

Although the CT scan is continuous, the UMAP embedding revealed **three broad anatomical regions**, each subdivided into several smaller clusters:

1. **Distal Tibia / Far Below the Joint (Slices 1–15)**
   These slices appeared on the far right of the UMAP plot and displayed simpler cortical geometry with minimal trabecular variation.
2. **Knee Joint Region (Slices 16–35)**
   This region formed a tall central structure. Slices 22 and 23 showed clear femoral condyles, tibial plateau, and joint spacing. UMAP separated this area into multiple small islands due to variation in cartilage thickness, condylar curvature, and joint surface geometry.
3. **Proximal Tibia / Lower Femur (Slices 36–72)**
   This region appeared on the left side of the embedding, with slice 45–46 illustrating metaphyseal transitions and slice 72 demonstrating increased femoral soft-tissue and cortical complexity.

Why More than Three Clusters Formed

UMAP produced **5–7 subclusters**, which is expected given:

- High sensitivity to local structural differences
- ResNet-50's ability to detect fine-grained tissue and bone features
- Anatomical complexity, especially around the femoral condyles and tibial plateau

Relation to Injury Dataset

Although the reviewed slices did not show classic fracture patterns, UMAP highlighted slices with **subtle feature deviations**—likely reflecting trabecular irregularities, density differences, or soft-tissue changes relevant to the larger injury dataset.

Discussion

The CNN-UMAP pipeline successfully separated slices into anatomically coherent groups without using labels. The presence of micro clusters within the joint region reflects the complex biomechanics and structure of the knee. This approach demonstrates how unsupervised learning can support tasks such as anomaly detection, slice prioritization, and structure-aware modeling pipelines.

Evaluation

The combined clinical and imaging analysis strengthened the predictive modeling results. Random Forest modeling identified strong clinical predictors of fractures and osteomyelitis, PCA revealed latent relationships connecting severity with bone outcomes, and UMAP supported these findings by visually distinguishing structural variations in CT anatomy. Together, these methods provide a comprehensive understanding of how burn injury characteristics correlate with deeper bone involvement.

Key Findings

This project revealed that burn severity, bone health descriptors, and hospital stay duration are the most significant predictors of internal bone complications. Material properties contributed meaningfully, and imaging embeddings demonstrated that even unsupervised models can capture anatomically relevant structure.

Limitations

Challenges included class imbalance, limited osteomyelitis cases, absence of longitudinal imaging, and variations in imaging quality. Despite these limitations, the multimodal framework proved effective and provides a strong foundation for further research.

References

[1] BurnCare Clinical Dataset, Meharry Medical College, 2025.

[2] KC Santosh et al., 'Medical Imaging Datasets,' GitHub, 2024.

[3] Khairi N., Predictive Modeling and Analysis Course Notes, 2025.

[4] Davis L., 'BurnCare Bone Health Prediction Midterm Presentation,' 2025.