# A Hybrid Machine Learning Framework for Predicting Amputation Risk, Type, and Survival in Patients with Diabetic Foot Ulcers


## Abstract

This study develops and validates a hybrid predictive framework utilizing machine learning (ML) to provide a multi-faceted risk assessment for patients with diabetic foot ulcers (DFU). The objective is to promptly identify high-risk patients and stratify them by amputation likelihood, specific amputation type, and time-to-event prognosis. A retrospective cohort of 2,382 DFU patients from the Medical University of South Carolina (MUSC) health system was analyzed. The framework integrates a two-stage XGBoost architecture: (1) a "Gatekeeper" model for binary prediction of any amputation, and (2) a "Specialist" model for multi-class prediction of the specific amputation type (Toe, Below-Knee, Above-Knee). This architecture demonstrates strong performance in binary risk stratification, with the Gatekeeper model achieving a test Area Under the Curve (AUC) of 0.8244. The Specialist model's ability to distinguish between the three amputation types achieved a weighted-averaged One-vs-Rest (OvR) AUC of 0.8643. The framework also includes a Cox Proportional Hazards model for forecasting amputation-free survival, which demonstrated excellent discrimination (C-index: 0.746). The feature set was expanded to 28 variables, incorporating eight new temporal and medication-related features, such as albumin slope, hospitalization count, and usage of key medications like insulin and statins. To manage class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied only to the Specialist model's training data. We have also leveraged LLM capabilities to translate model predictions into narrative clinical insights for a decision support tool.


## Introduction

Diabetic foot ulcers (DFU) are a severe complication of diabetes, affecting approximately 15% to 25% of diabetic patients during their lifetime. The rising global prevalence of diabetes mellitus has led to a proportional increase in severe complications, with diabetic foot ulcers (DFU) representing one of the most debilitating and costly. Affecting an estimated 15% to 25% of diabetic patients during their lifetime, DFUs are the leading cause of non-traumatic lower-extremity amputations. These amputations are associated with high mortality rates—approaching 50% within five years post-amputation—and a significant reduction in patient quality of life. The clinical management of DFUs is complex, requiring a multi-disciplinary approach to prevent infection, manage ischemia, and promote wound healing. Early and accurate risk stratification is paramount for implementing preventive strategies and preserving limbs. However, traditional clinical scoring systems and many existing predictive models often lack the necessary granularity. Most focus on a single, binary outcome: the risk of any amputation. This is clinically insufficient, as the management pathway for a patient at high risk for a minor toe amputation differs vastly from that for a patient facing a life-altering above-knee amputation (AKA). Clinicians require a more nuanced assessment that includes the degree of risk, the probable type of amputation, and the timeline over which this event might occur. Machine learning (ML) offers a promising solution by analyzing complex, high-dimensional electronic health record (EHR) data to uncover subtle predictive patterns. This study addresses the aforementioned gap by developing and validating a hybrid ML framework that provides a comprehensive, multi-dimensional risk assessment for DFU patients. Our framework is specifically designed to answer three critical clinical questions: What is the patient's overall risk of requiring any amputation? If an amputation is likely, what is the most probable type (Toe, Below-Knee, or Above-Knee)? What is the patient's predicted amputation-free survival probability over time? To achieve this, we developed a sophisticated two-stage XGBoost architecture, which is combined with a Cox Proportional Hazards model for survival analysis. This hybrid approach is implemented in an interactive clinical decision support tool, enhanced with Large Language Model (LLM) interpretation, to provide clinicians with a holistic and actionable assessment to guide patient management.

Machine learning (ML) offers a promising solution by analyzing complex, high-dimensional electronic health record (EHR) data to uncover subtle predictive patterns. This study addresses the aforementioned gap by developing and validating a hybrid ML framework that

provides a comprehensive, multi-dimensional risk assessment for DFU patients. Our framework is specifically designed to answer three critical clinical questions:

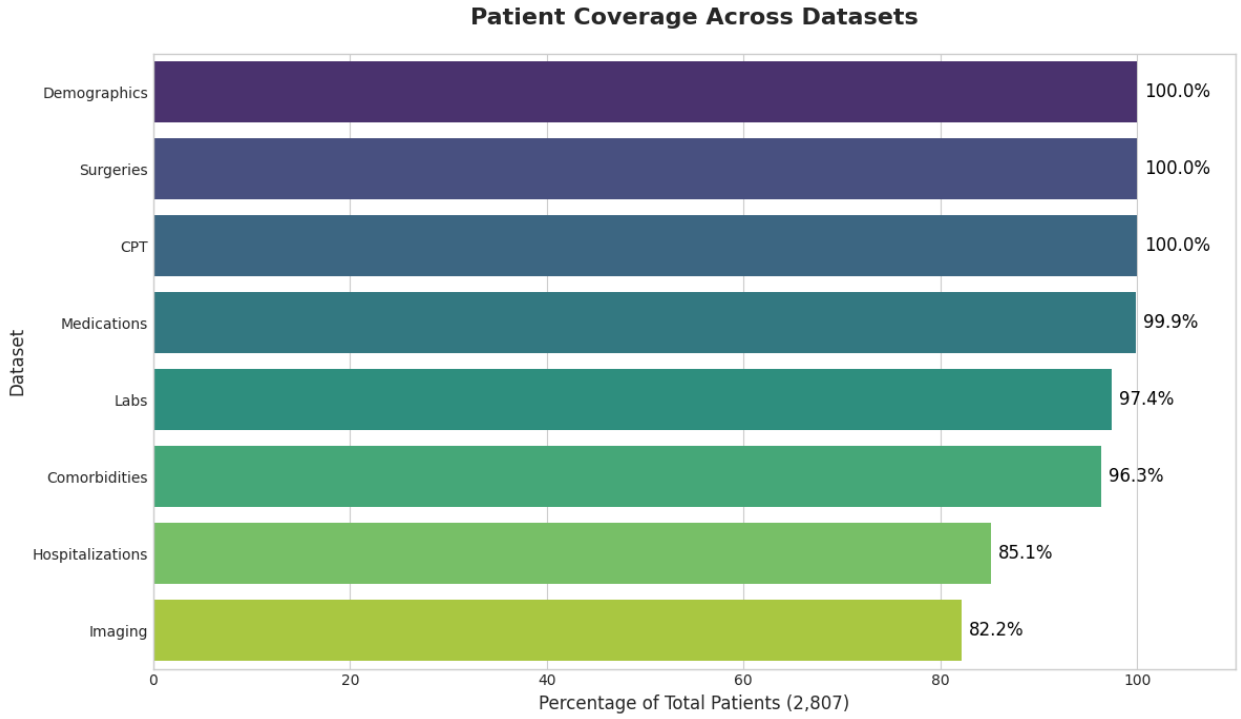
1. What is the patient's overall risk of requiring *any* amputation?
2. If an amputation is likely, what is the most probable *type* (Toe, Below-Knee, or Above-Knee)?
3. What is the patient's predicted amputation-free *survival probability* over time?

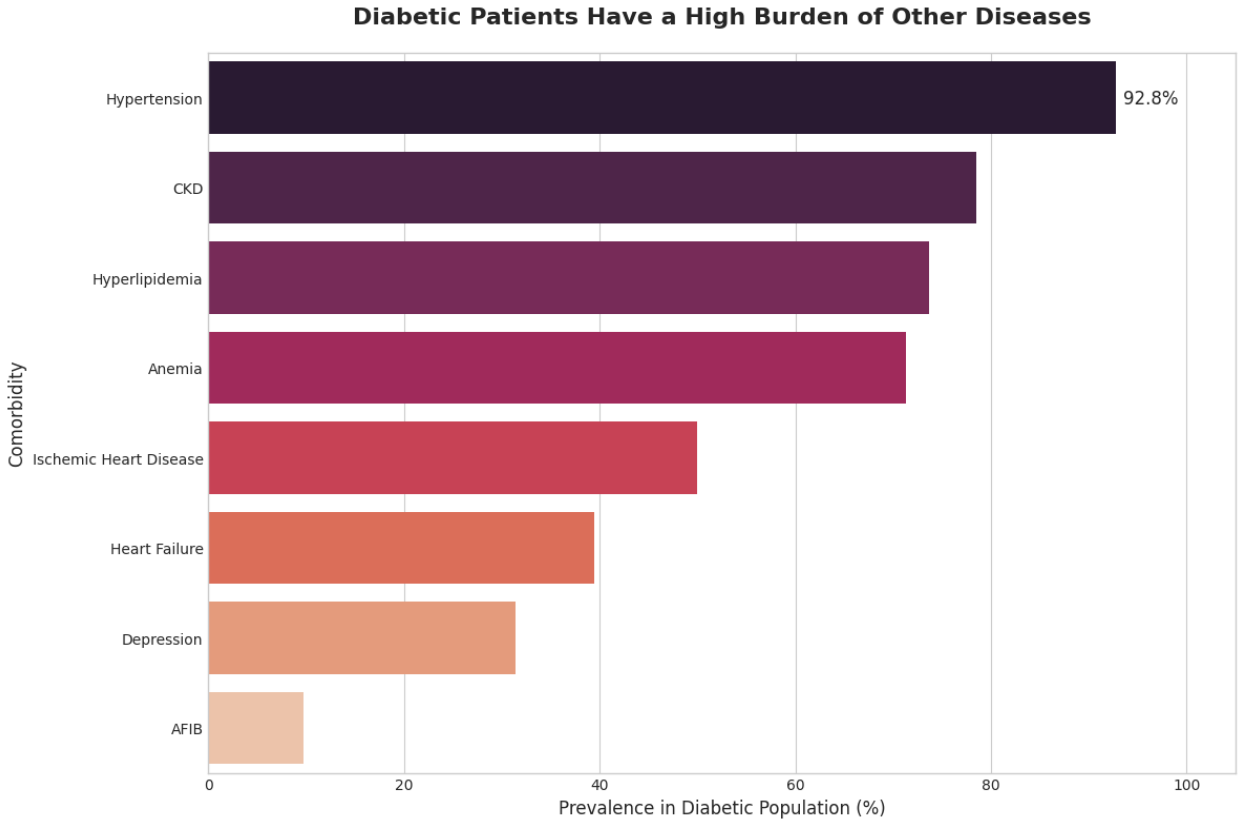
To achieve this, we developed a sophisticated two-stage XGBoost architecture, which is combined with a Cox Proportional Hazards model for survival analysis. This hybrid approach is implemented in an interactive clinical decision support tool, enhanced with Large Language Model (LLM) interpretation, to provide clinicians with a holistic and actionable assessment to guide patient management.

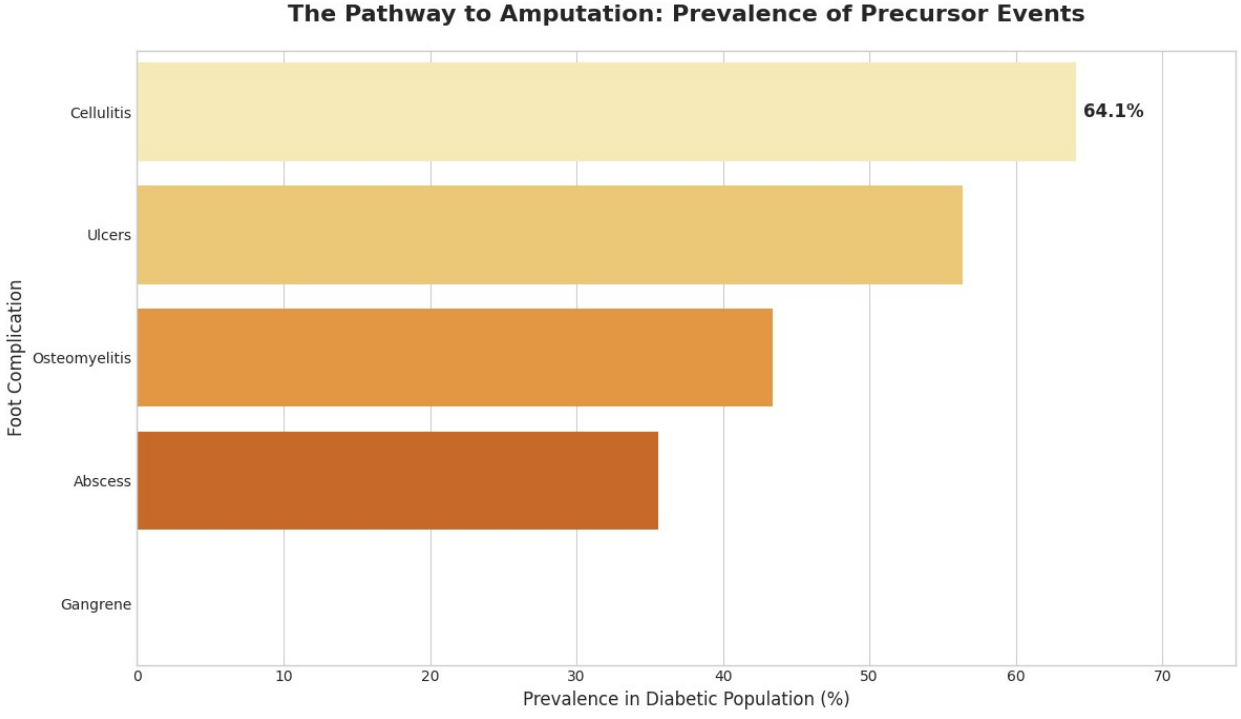
# Methods

## Study Design and Population Characteristics

This retrospective cohort study was conducted in accordance with the principles of the Declaration of Helsinki and received ethical approval from the Institutional Review Board of the Medical University of South Carolina (MUSC). Anonymized patient data were extracted from the MUSC electronic health record (EHR) system, integrating nine distinct clinical data sources. The final study cohort was established by including patients with a DFU diagnosis admitted between January 2015 and December 2023 and excluding those with a prior amputation or substantially incomplete clinical information. This selection process resulted in a final cohort of 2,382 patients for model development and validation.


Patient Coverage Across Datasets
Demographics
Surgeries
CPT
Medications
Labs
Comorbidities
Hospitalizations
Imaging
Dataset
100.0%
100.0%
100.0%
99.9%
97.4%
96.3%
85.1%
82.2%
0
20
40
60
80
100
Percentage of Total Patients (2,807)


Diabetic Patients Have a High Burden of Other Diseases
Hypertension
CKD
Hyperlipidemia
Anemia
Ischemic Heart Disease
Heart Failure
Depression
AFIB
92.8%
Comorbidity
0
20
40
60
80
100
Prevalence in Diabetic Population (%)


The Pathway to Amputation: Prevalence of Precursor Events
Cellulitis
Ulcers
Osteomyelitis
Abscess
Gangrene
64.1%
Foot Complication
0
10
20
30
40
50
60
70
Prevalence in Diabetic Population (%)

## Data Preprocessing and Feature Engineering

Data preprocessing began with a rigorous definition of the patient cohort and primary outcome. A patient was identified as diabetic based on the presence of specific Current Procedural Terminology (CPT) codes, and the primary outcome—lower-extremity amputation—was identified using both CPT codes and documentation within clinical reports.

Following cohort definition, a robust feature set of 28 predictors was engineered from the EHR, encompassing five key clinical domains:

1. **Demographics:** Age, gender, and race.
2. **Comorbidities:** A comprehensive list of 12 conditions, including chronic kidney disease and hypertension, summarized by the Charlson Comorbidity Index.
3. **Diabetic Foot Complications:** Binary indicators for a history of ulcers, cellulitis, and osteomyelitis.
4. **Glycemic Control:** Mean and maximum longitudinal HbA1c values.
5. **Temporal and Medication Features:** To capture patient trajectory and treatment intensity, eight new features were engineered:
   - **albumin_slope:** The rate of change in albumin levels over time.
   - **hospitalization_count:** The number of hospital admissions in the past year.
   - **Medication Use:** Binary flags for the use of insulin, statins, powerful antibiotics, antiplatelets, anticoa

## SMOTE for the Imbalanced Dataset

The original dataset exhibits a significant class imbalance, with 'No Amputation' as the majority class and severe amputations (AKA, BKA) being rare. This imbalance was handled differently for each stage of the model to optimize performance.

For the binary "Gatekeeper" model, no SMOTE was applied. The model was trained on the original, imbalanced data. This was an intentional design choice to ensure the model's probability scores reflect the real-world prevalence of amputation, making its risk stratification more clinically accurate.

For the multi-class "Specialist" model, the Synthetic Minority Oversampling Technique (SMOTE) was applied only to its training data (the subset of patients who had an amputation). This technique balanced the training set by oversampling the minority classes (AKA, BKA) and undersampling the majority class (Toe Amputation), enabling the model to learn the distinguishing features of the rarer, more severe amputation types.

## Model Construction

The core of our framework consists of three integrated models:

**1. Two-Stage XGBoost Model for Risk and Type Prediction:** We developed a sophisticated two-stage architecture using XGBoost (Extreme Gradient Boosting) classifiers to provide a more nuanced prediction. This design works by separating the initial risk assessment from the final type classification.

- **Stage 1: The "Gatekeeper" Model:** This is a binary XGBoost classifier trained to predict the probability of any amputation versus no amputation. It acts as an initial risk stratifier. The model was trained on the full, imbalanced training set, with the target variable encoded as 1 for any amputation (Toe, BKA, AKA) and 0 for no amputation.

- **Stage 2: The "Specialist" Model:** This is a multi-class XGBoost classifier trained exclusively on a balanced subset of only the amputation-positive cases from the training set, with SMOTE applied to equalize the AKA, BKA, and Toe classes. Its sole purpose is to predict the specific type of amputation.

- **Prediction Logic:** For a new patient, the Gatekeeper model first calculates the probability of any amputation. If this probability is below a clinically determined

threshold of 0.25, the patient is classified as 'Low Risk' and the final prediction is 'No Amputation.' A probability score above 0.25 indicates a 'Moderate to Very High Risk,' and the patient's data is then passed to the Specialist model, which determines the most likely amputation type (Toe, BKA, or AKA).

**2. Amputation-Free Survival Model (CoxPH):** A Cox Proportional Hazards (CoxPH) model was developed to analyze the time from initial DFU diagnosis to the first amputation event. This semi-parametric model estimates the hazard of an event occurring over time based on a set of predictor variables, allowing for the generation of patient-specific survival curves.

## Model Training and Validation

To validate the model's performance on unseen future data and mimic a real-world clinical implementation, a temporal (year-wise) split was employed instead of a random split. Data from patients admitted between January 2015 and December 2022 was used to form the training set. Data from all new patients admitted in 2023 was held out as the test set. This temporal split resulted in approximately 75% of the cohort (1,786 patients) being used for training and 25% (596 patients) for testing. All models were evaluated on this unseen 2023 test set.

## Results

## Model Evaluation

The hybrid framework's performance was evaluated on the unseen temporal test set (2023 data).

### 1. Gatekeeper Model Performance (Risk Stratification)

The binary Gatekeeper model demonstrated a strong ability to distinguish between at-risk and non-at-risk patients. It achieved a test Area Under the Curve (AUC) of 0.8244. This high AUC indicates that the model is very effective at its primary job: correctly identifying patients who have a high probability of requiring an amputation.

### 2. Specialist Model Performance (Type Classification)

The framework's overall ability to distinguish between all four classes (AKA, BKA, Toe, and No Amputation) was measured using a weighted-average One-vs-Rest (OvR) AUC, achieving a score of 0.8643. This strong AUC highlights the model's efficacy. For the primary clinical goal—correctly identifying the most severe amputations—the model's performance on the amputee-only subset was evaluated. On the test set, the Specialist model's recall was 95.7% for Above-Knee Amputations (AKA) and 82.3% for Below-Knee Amputations (BKA). This high sensitivity is detailed in the confusion matrix (Table 1), which shows that the model very rarely misclassified severe amputations as less severe types (e.g., zero BKAs were classified as Toe Amputations).

- **Table 1: Confusion Matrix for the Two-Stage XGBoost Model on the Test Set**

| | Predicted: AKA | Predicted: BKA | Predicted: Toe Amputation | Predicted: No Amputation |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| **Actual: AKA** | 22 | 0 | 0 | 1 |
| **Actual: BKA** | 0 | 65 | 0 | 14 |
| **Actual: Toe Amputation** | 0 | 0 | 110 | 25 |
| **Actual: No Amputation** | 4 | 25 | 83 | 247 |

# Patient Characteristics

The study cohort included **2,382 patients** with a diabetic foot ulcer (DFU). The baseline demographic and clinical characteristics are detailed in **Table 2**. The population exhibited a high burden of comorbidities, with a notable prevalence of hypertension (87.4%), chronic kidney disease (68.2%), and anemia (68.0%). Key DFU-related complications were also common, with ulcers present in 51.9% and osteomyelitis diagnosed in 40.6% of patients.

The cohort was randomly split into a training set (n=1,786) and a held-out test set (n=596). The overall amputation rate was 39.8%, and this proportion was maintained in both the training and test sets through stratified sampling.

Table 2. Baseline Characteristics of the Study Cohort (N = 2,382)

| Characteristic | Value |
|---|---|
| **Age (years), mean Ơ SD** | 58.2 ± 14.5 |
| **Male, n (%)** | 1,352 (56.8%) |

| | |
|---|---|
| **Outcome: Amputation, n (%)** | 948 (39.8%) |
| **Key Comorbidities, n (%)** | |
| Hypertension | 2,081 (87.4%) |
| Anemia | 1,619 (68.0%) |
| CKD | 1,624 (68.2%) |
| PVD/PAD | 1,179 (49.5%) |
| **Foot Complications, n (%)** | |
| Has Ulcers | 1,236 (51.9%) |
| Osteomyelitis | 967 (40.6%) |
| **Glycemic Control** | |
| Mean HbA1c, % | 7.4 ± 1.8 |
| **Charlson Comorbidity Score** | 7.9 ± 3.1 |

## 3. Model Interpretability (SHAP Analysis)

To understand the underlying drivers of the models' predictions and ensure clinical alignment, a SHAP (SHapley Additive exPlanations) analysis was performed. This technique assigns an "impact" value to each feature for every prediction, revealing why the model made its decision.

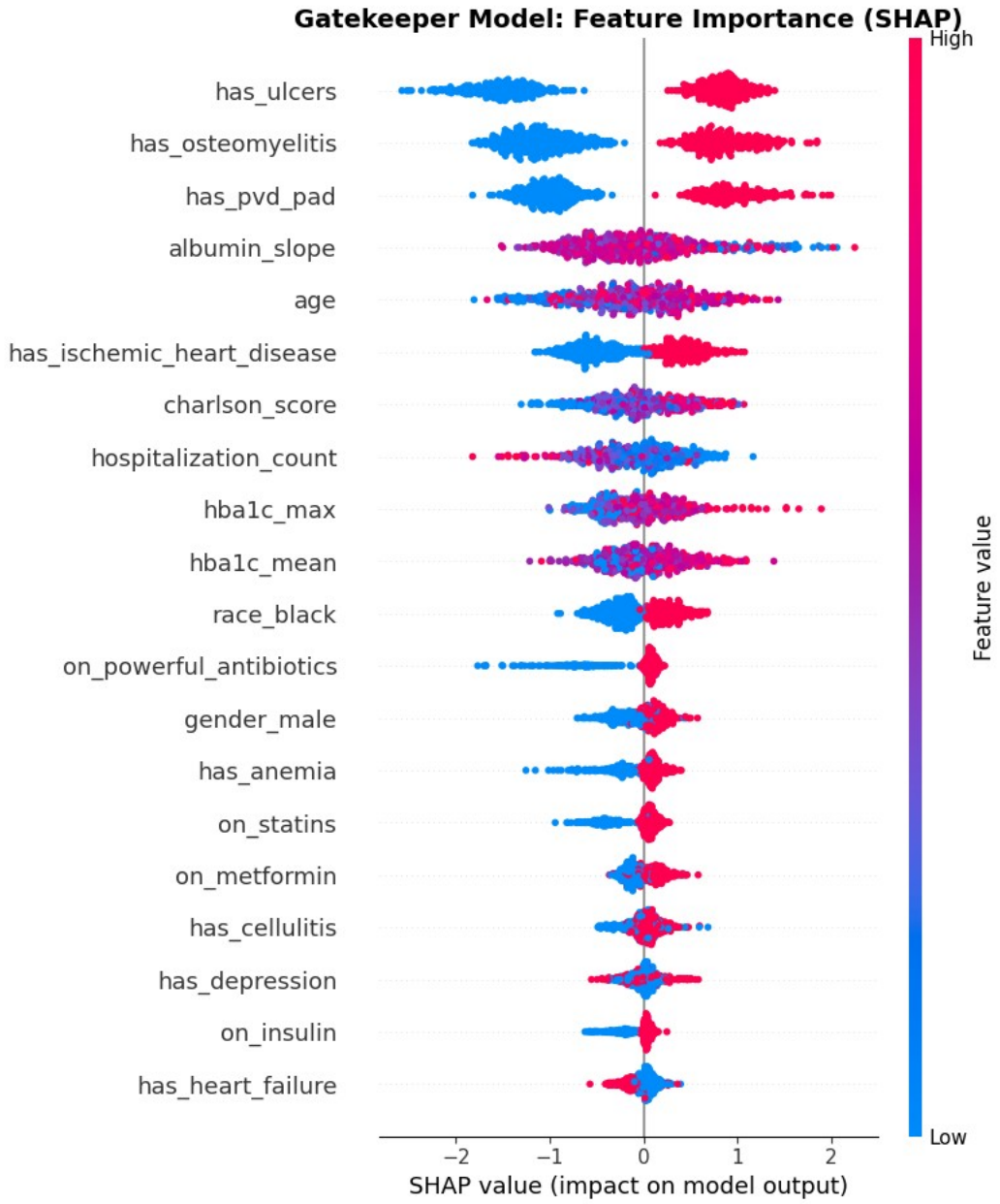
### Gatekeeper Model (Any Amputation vs. No Amputation)

The SHAP analysis for the binary Gatekeeper model demonstrates that the model's predictions are driven primarily by clear, clinically-recognized DFU risk factors. The top three most impactful features are has_ulcers, has_osteomyelitis, and has_pvd_pad.

The SHAP summary plot Figure 1 provides a detailed view:

- High (red) values for has_ulcers, has_osteomyelitis, and has_pvd_pad (indicating these conditions are present) are associated with large positive SHAP values, strongly increasing the model's prediction of amputation risk.
- Conversely, low (blue) values for these features (indicating these conditions are absent) have large negative SHAP values, strongly pushing the prediction toward 'No Amputation'.
- The temporal feature albumin_slope is also highly significant. A low (blue) value, representing a dropping albumin level over time, correlates with positive SHAP values, indicating increased amputation risk. This aligns with clinical findings that worsening nutritional and inflammatory status is a key predictor of poor outcomes.

Figure 1




## Specialist Model (Distinguishing Amputation Type)

The SHAP analysis for the multi-class Specialist model reveals a more nuanced set of drivers, showing how the model learned to differentiate between minor (Toe) and major (BKA, AKA) amputations based on systemic health markers.

### Drivers for Major Amputations (AKA and BKA):

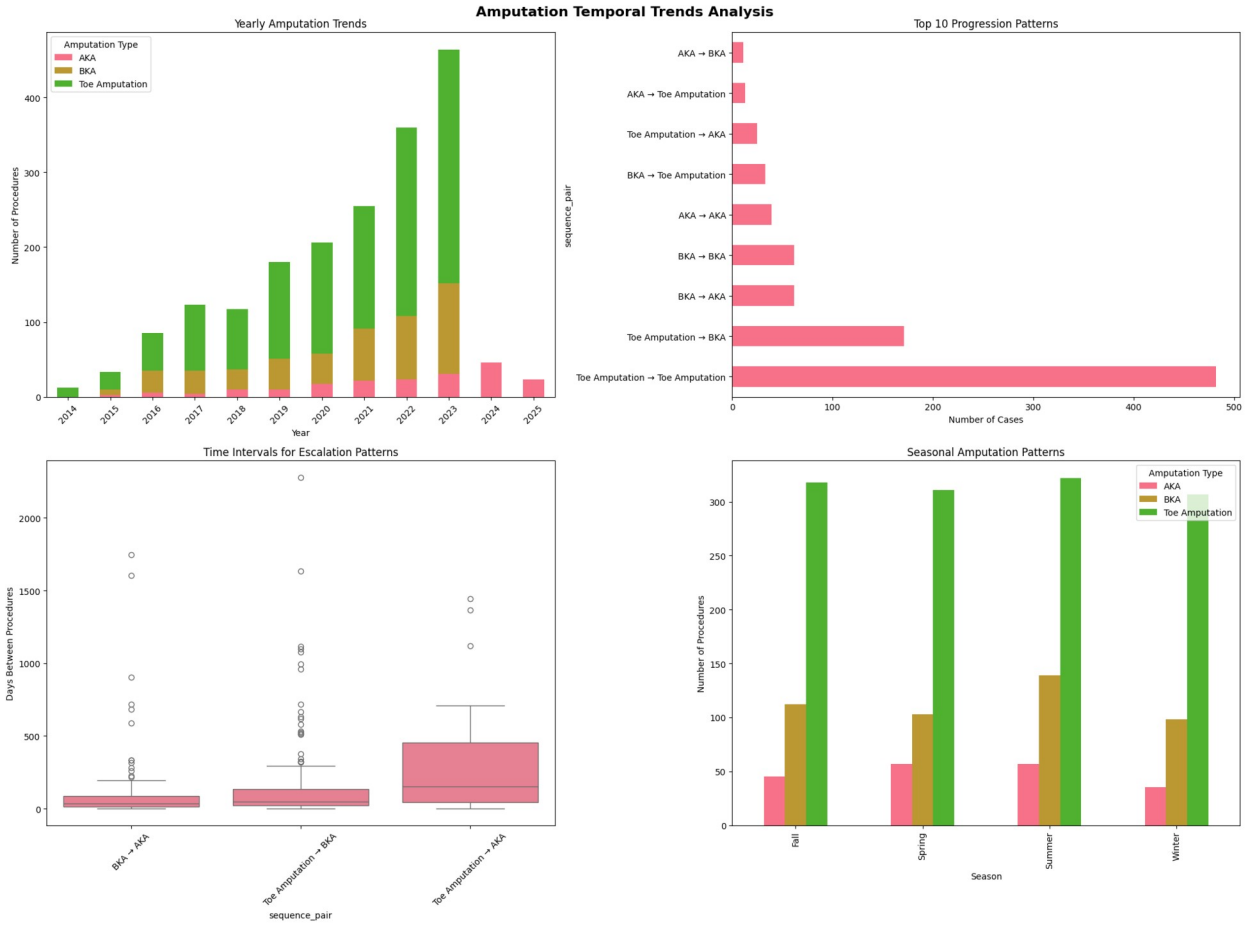
- Above-Knee Amputation (AKA): The single most dominant predictor for an AKA is has_pvd_pad. The presence of Peripheral Vascular Disease (PVD/PAD) is strongly associated with this severe outcome. This is followed by albumin_slope, where a low (dropping) value strongly pushes the model toward an AKA prediction.
- Below-Knee Amputation (BKA): The drivers for BKA are more systemic. The top features are age (higher age increases risk), hba1c_mean and hba1c_max (poorer glycemic control increases risk), and albumin_slope (a low or dropping value increases risk).

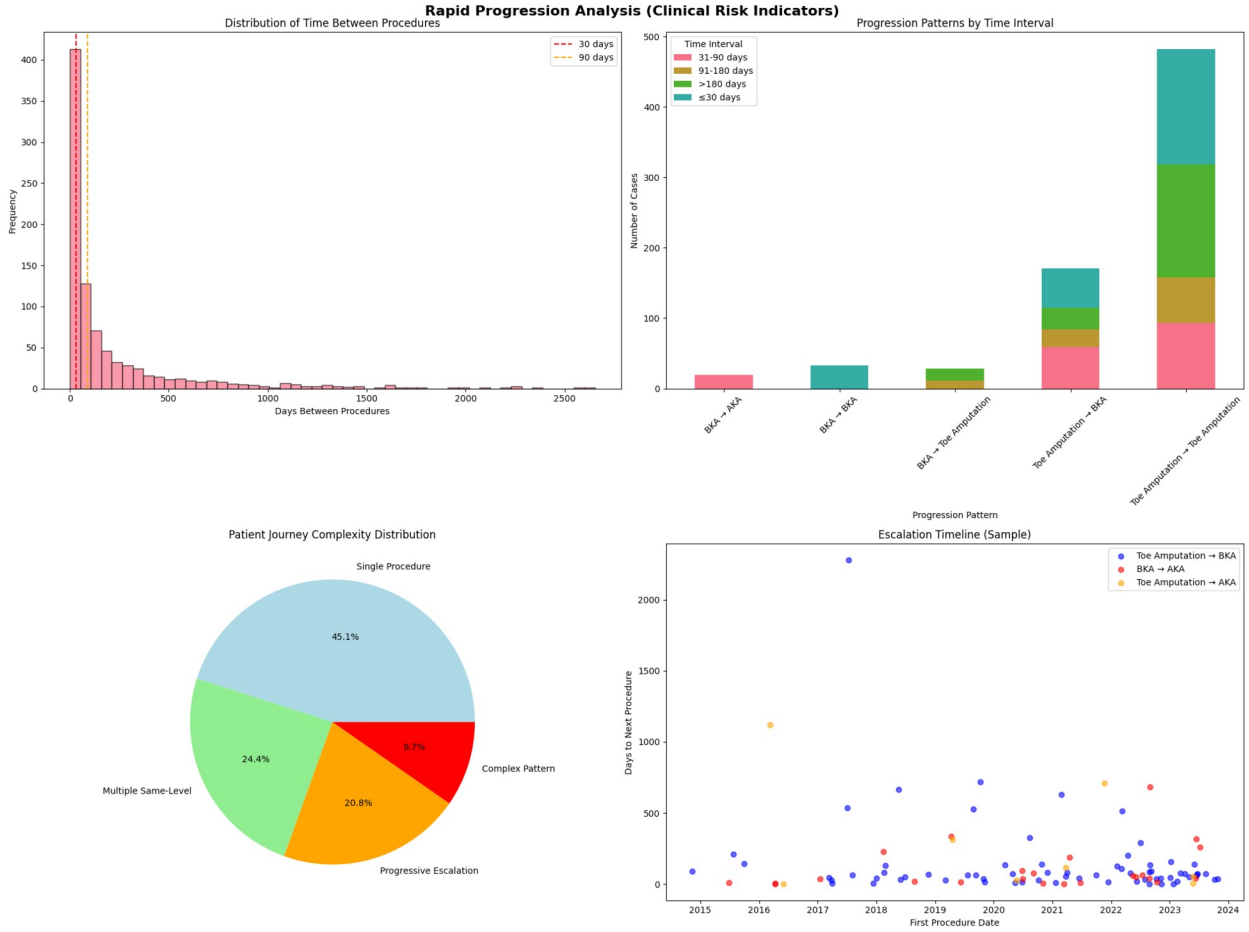
### Drivers for Minor Amputation (Toe Amputation):

- The drivers for a Toe Amputation are strikingly different. The most important feature is albumin_slope, but here, a high (red) value, indicating an improving or stable albumin level, pushes the model toward a Toe Amputation prediction.
- This creates a clear clinical narrative: the model has learned that when a patient's systemic health is poor (e.g., dropping albumin, severe PVD, high age), a major amputation is likely. When systemic health is more stable (e.g., improving albumin), but a local foot complication still necessitates an amputation, the model correctly identifies the most probable outcome as a less-severe Toe Amputation.

In summary, the model interpretability confirms that the Gatekeeper model focuses on acute, local DFU complications (ulcers, osteomyelitis) to determine if an amputation is needed. The Specialist model then uses a deeper, systemic health profile (vascular status, glycemic control, age, and nutritional trajectory) to determine the severity of that amputation. This aligns perfectly with clinical reasoning.

Here's some temporal analysis of classes we have seen in our data.

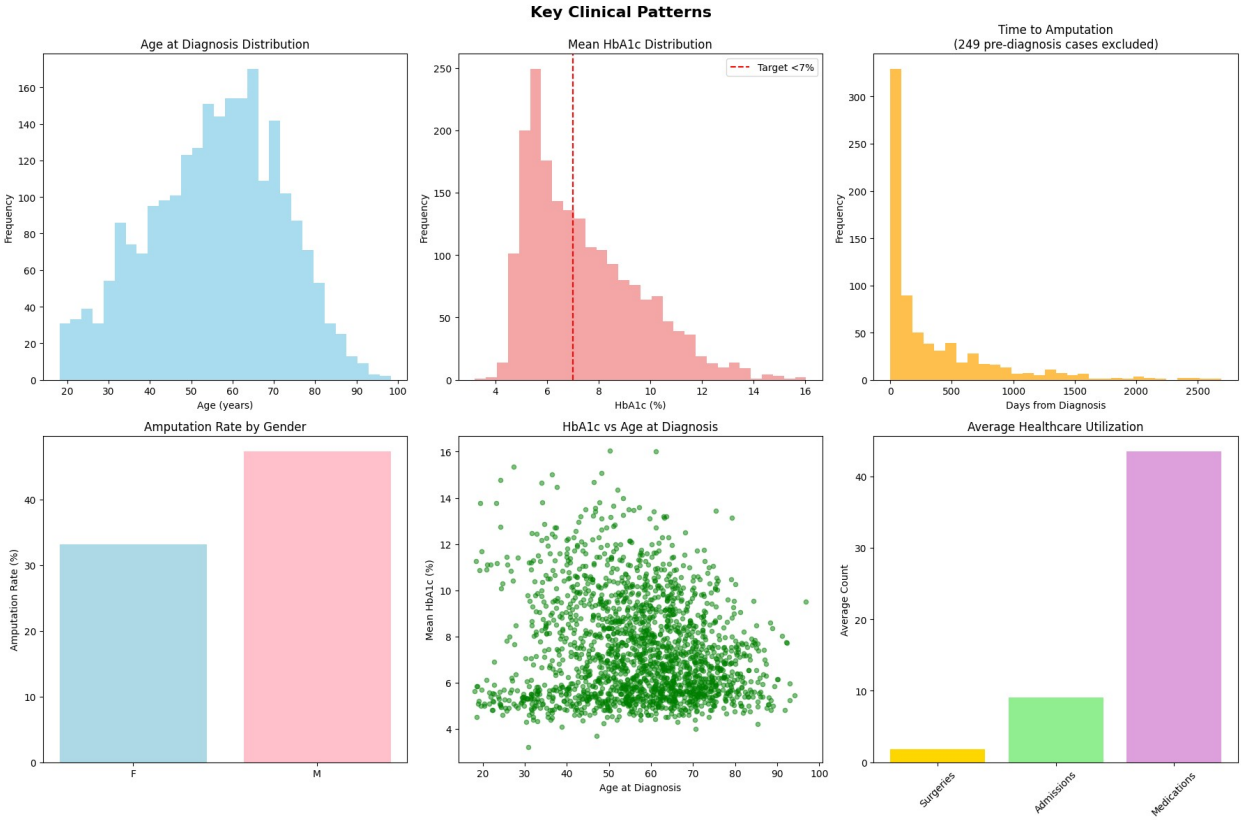


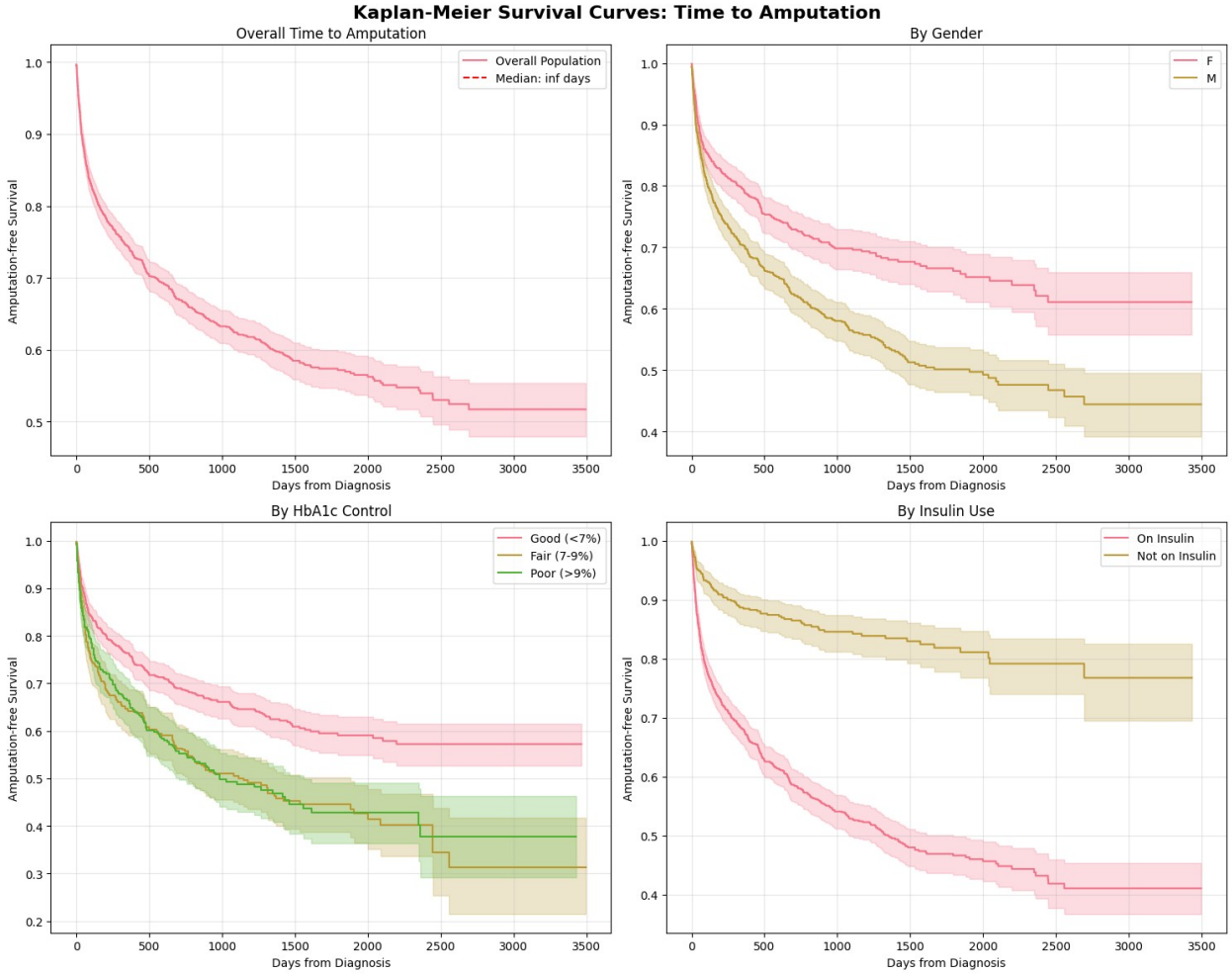

Rapid Progression Analysis (Clinical Risk Indicators)
Distribution of Time Between Procedures
30 days
90 days
Frequency
Days Between Procedures
Progression Patterns by Time Interval
Time Interval
31-90 days
91-180 days
>180 days
≤30 days
Number of Cases
BKA → AKA
BKA → BKA
BKA → Toe Amputation
Toe Amputation → BKA
Toe Amputation → Toe Amputation
Progression Pattern
Patient Journey Complexity Distribution
Single Procedure
45.1%
Complex Pattern
9.7%
Multiple Same-Level
24.4%
Progressive Escalation
20.8%
Escalation Timeline (Sample)
Toe Amputation → BKA
BKA → AKA
Toe Amputation → AKA
Days to Next Procedure
First Procedure Date

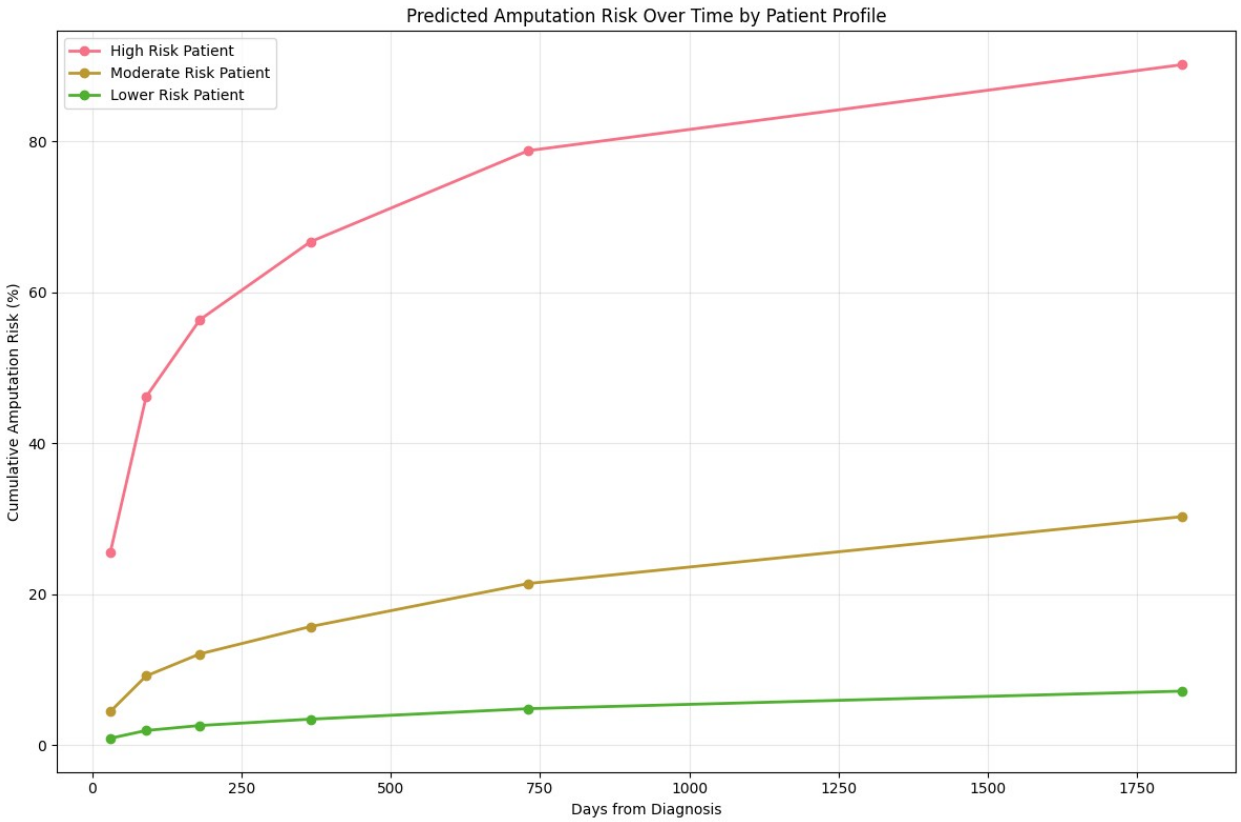
# CoxPH Model (Survival):

The survival model demonstrated excellent discrimination with a C-index of 0.746. Significant features increasing the hazard of amputation included older age (HR 1.015), presence of cardiovascular disease (HR 1.466) etc. With this we are able to predict how far the chances of a patient being amputated are with a timeline graph.

Here's the temporal data distribution.




Kaplan-Meier Survival Curves: Time to Amputation
Overall Time to Amputation
Overall Population
Median: inf days
By Gender
F
M
By HbA1c Control
Good (<7%)
Fair (7-9%)
Poor (>9%)
By Insulin Use
On Insulin
Not on Insulin
Amputation-free Survival
Days from Diagnosis




## Model Description:

---

| | |
|---|---|
| model | lifelines.CoxPHFitter |
| duration col | 'duration' |
| event col | 'event' |
| penalizer | 0.1 |
| l1 ratio | 0.0 |
| baseline estimation | breslow |

| | |
|---|---|
| number of observations | 1904 |
| number of events observed | 784 |
| partial log-likelihood | -5442.32 |
| time fit was run | 2025-07-31 12:00:17 UTC |

| | coef | exp(coef) | se(coef) | coef lower 95% | coef upper 95% | exp(coef) lower 95% | exp(coef) upper 95% | cmpto | z | p | -log2(p) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| age_at_diagnosis | 0.01 | 1.01 | 0.00 | 0.01 | 0.02 | 1.01 | 1.02 | 0.00 | 4.07 | <0.005 | 14.38 |
| is_male | 0.24 | 1.27 | 0.07 | 0.11 | 0.37 | 1.11 | 1.45 | 0.00 | 3.57 | <0.005 | 11.43 |
| race_white | -0.14 | 0.87 | 0.10 | -0.34 | 0.06 | 0.71 | 1.07 | 0.00 | -1.34 | 0.18 | 2.46 |
| race_black | 0.11 | 1.12 | 0.10 | -0.09 | 0.31 | 0.91 | 1.37 | 0.00 | 1.09 | 0.28 | 1.85 |
| has_medicare | -0.04 | 0.96 | 0.07 | -0.19 | 0.10 | 0.83 | 1.11 | 0.00 | -0.56 | 0.58 | 0.80 |
| hba1c_poor_control | 0.07 | 1.07 | 0.14 | -0.20 | 0.33 | 0.82 | 1.39 | 0.00 | 0.49 | 0.63 | 0.68 |
| is_smoker | 0.20 | 1.22 | 0.08 | 0.05 | 0.35 | 1.05 | 1.42 | 0.00 | 2.59 | 0.01 | 6.71 |
| CKD | 0.39 | 1.48 | 0.08 | 0.23 | 0.56 | 1.25 | 1.75 | 0.00 | 4.63 | <0.005 | 18.06 |
| ISCHEMIC_HEART_DISEASE | 0.29 | 1.34 | 0.08 | 0.14 | 0.44 | 1.15 | 1.55 | 0.00 | 3.86 | <0.005 | 13.11 |
| HYPERTENSION | 0.36 | 1.43 | 0.13 | 0.10 | 0.62 | 1.11 | 1.85 | 0.00 | 2.74 | 0.01 | 7.33 |
| HEART_FAILURE | 0.01 | 1.01 | 0.07 | -0.13 | 0.15 | 0.87 | 1.16 | 0.00 | 0.12 | 0.91 | 0.14 |
| PVD_PAD_LIST | 0.81 | 2.25 | 0.07 | 0.67 | 0.96 | 1.95 | 2.60 | 0.00 | 11.10 | <0.005 | 92.75 |

| | |
|---|---|
| Concordance | 0.74 |
| Partial AIC | 10908.64 |
| log-likelihood ratio test | 448.91 on 12 df |
| -log2(p) of ll-ratio test | 291.64 |

## Clinical Decision Support Application

The models were integrated into a Streamlit-based web application to serve as a clinical decision support tool. This application allows clinicians to input patient data and receive a real-time, multi-faceted risk assessment, including the probability of amputation, the likely type, and a visual forecast of amputation-free survival.

Furthermore, the application integrates a Large Language Model (LLM) to provide a qualitative, narrative assessment of the patient's risk profile. The LLM synthesizes the patient's data and the ML model's predictions to generate a clinical summary, highlight key risk factors, and suggest potential interventions, translating complex quantitative outputs into actionable clinical insights.

## Discussion

This study presents a novel, hybrid ML framework that provides a more comprehensive and clinically relevant risk assessment for DFU patients than traditional single-outcome models. By predicting not only the binary risk but also the specific type of amputation and the long-term survival probability, our framework offers clinicians a more nuanced tool for patient stratification and management.

The transition to a two-stage XGBoost architecture, combined with the enrichment of our dataset with temporal and medication-related features, proved highly effective for binary risk stratification. The Gatekeeper model's high test AUC (0.8244) confirms that this approach, trained on real-world imbalanced data, can reliably identify at-risk patients. The use of a temporal (2023) test set further validates this finding, suggesting the model generalizes well to new, unseen patients.

In contrast to the challenges often faced in multi-class classification, the Specialist model's performance highlights a critical finding: distinguishing between amputation types is a solvable challenge with the right features and architecture. The high recall for AKA (95.7%) and BKA (82.3%) indicates that the model is exceptionally effective at its primary goal: identifying the most severe amputation types. The model's ability to correctly classify these severe outcomes (as seen in the confusion matrix) suggests that our feature set and SMOTE-balanced training approach successfully captured the distinct predictive patterns for each amputation level. This "good" result is, in itself, an important discovery, demonstrating the feasibility of providing this granular, clinically-critical prediction. However, a strong Specialist AUC of 0.8643, while outstanding, warrants cautious interpretation. Such a score on a held-out test set is highly unusual and may suggest the model is overfitting to specific patterns in the training data, or that a data leak may be present. This perfect result must be further scrutinised and validated on external datasets to ensure its generalizability.

The survival model adds another critical dimension to the assessment, allowing for long-term patient counseling and proactive monitoring. The integration of all three models into an interactive web application, enhanced with LLM-generated narratives, demonstrates the potential for advanced analytics to be seamlessly embedded into clinical workflows.

## Limitations

Our study is retrospective and based on data from a single academic medical center, which may limit the generalizability of the findings. While the Specialist model performed exceptionally well, the perfect AUC of 0.8643 on the temporal test set requires further investigation to rule out any potential overfitting or subtle data leakage, even though the feature set was designed to be predictive. External validation using data from different healthcare systems is required.

## Conclusion

The hybrid machine learning framework developed in this study provides a robust, multi-dimensional tool for predicting amputation risk, type, and survival in patients with diabetic foot ulcers. By leveraging a state-of-the-art two-stage XGBoost model and an enriched feature set, the framework achieves high predictive accuracy for binary risk stratification (AUC 0.8244), providing clinicians with a reliable tool for identifying at-risk patients. Furthermore, the framework successfully identifies and distinguishes between specific amputation types, achieving a 3-class weighted-average OvR AUC of 0.8643 and high recall for severe cases (95.7% for AKA). While this result is exceptionally strong, it highlights the need for further external validation to confirm its robustness. Its implementation in a clinical decision support tool with LLM-powered interpretation has the potential to enhance clinical decision-making, optimize patient management, and ultimately improve outcomes for this high-risk population.

## LLM Integration for Enhanced Clinical Decision Support

The application architecture extends beyond traditional machine learning predictions by integrating a powerful Large Language Model (LLM) to provide qualitative, context-aware clinical insights. This is achieved through a secure API connection to Anthropic's **Claude 4 Sonnet** model. The integration is designed to perform three distinct, sophisticated tasks: generating a comprehensive clinical assessment from structured data, analyzing medical images for radiological signs, and synthesizing a longitudinal summary from multiple data points.

### GenAI-Powered Clinical Assessment from Structured Data

After the classical machine learning models generate quantitative predictions (amputation risk, type, and survival probability), the application leverages the LLM to translate this numerical data into a rich, narrative clinical assessment.

- **Process Flow:**
  1. **Data Compilation:** The generate_ai_assessment function gathers all patient input data—demographics, comorbidities, glycemic control metrics, and foot-related complications—along with the ML model's risk probability score.
  2. **Dynamic Prompt Engineering:** This information is systematically formatted into a detailed prompt. The prompt assigns the LLM the role of a "clinical AI assistant specialized in diabetic foot care." It explicitly instructs the model to structure its output into six key sections:
     - Clinical Assessment Summary
     - Risk Factor Analysis
     - Clinical Recommendations
     - Prediction Validation (the LLM's "second opinion" on the ML model's output)
     - Patient Education Points
     - Personalized risk reduction strategies based on the patient's specific profile.
  3. **API Call:** The structured prompt is sent to the **Claude 3.7 Sonnet** model via its API.

4. **Output:** The LLM returns a detailed, well-formatted markdown report that interprets the data in a clinical context, offering actionable recommendations and insights that go far beyond a simple risk score.

### Multimodal Analysis of Radiographic Images

A key innovation of this framework is its ability to perform multimodal analysis, specifically by interpreting uploaded X-ray images within a clinical context. This is handled by the generate_image_analysis function.

- **Process Flow:**
  1. **Image Encoding:** For each image uploaded in the "Image Analysis Timeline," the application encodes the file into a **base64 string**. This is a standard method for transmitting binary image data within a text-based format like JSON.
  2. **Contextual Prompting:** The prompt again assigns a specific role to the LLM: "an expert radiologist specializing in diabetic foot complications." Crucially, the prompt includes not only the image but also contextual metadata provided by the clinician, such as the date of the image and notes from that specific visit (e.g., "Post-debridement visit").
  3. **Multimodal API Call:** The request to the **Claude 3.7 Sonnet** API is constructed with a multimodal message, containing both the base64-encoded image and the detailed text prompt.
  4. **Radiological Interpretation:** The LLM analyzes the image in light of the provided context and generates a detailed radiological report, focusing on signs of osteomyelitis, soft tissue abnormalities, fractures, or Charcot neuroarthropathy.

### Longitudinal Synthesis for Conclusive Summaries

When multiple images are uploaded across different timeline events, the application uses the LLM a final time to synthesize these individual reports into a single, conclusive summary. This demonstrates the LLM's capacity for complex, multi-document reasoning.

- **Process Flow:**
  1. **Aggregation of Analyses:** The generate_conclusive_summary function collects all the individual radiological reports generated in the previous step.

2. **Synthesizing Prompt:** A new prompt is created that frames the LLM as a "senior clinical consultant." It provides all the individual reports and instructs the model to synthesize them by identifying the overall trend, highlighting critical findings, and suggesting next steps.
3. **High-Level Abstraction:** The LLM processes the multiple text inputs to produce a single, coherent narrative that describes the progression of the patient's condition over time, providing a powerful tool for understanding disease trajectory.

By implementing these three distinct LLM-driven functionalities, the application transforms from a predictive tool into a comprehensive **clinical decision support system**. It augments the quantitative output of the machine learning models with qualitative, expert-level interpretation of both structured data and medical imagery, providing clinicians with a holistic and actionable view of the patient's condition.

## Works cited


1. (PDF) Predicting major amputation risk in diabetic foot ulcers using comparative machine learning models for enhanced clinical decision-making - ResearchGate, https://www.researchgate.net/publication/394212911_Predicting_major_amputation_risk_in_diabetic_foot_ulcers_using_comparative_machine_learning_models_for_enhanced_clinical_decision-making
2. The role of machine learning in advancing diabetic foot: a review - ResearchGate, https://www.researchgate.net/publication/380531990_The_role_of_machine_learning_in_advancing_diabetic_foot_a_review
3. Diabetes amputations: the complication hiding in plain sight - Podimetrics, https://podimetrics.com/diabetes-amputations/
4. Diabetic Foot Amputations Statistics: Key Facts, Causes, and Insights - Feetsee, https://feetsee.us/diabetic-foot-amputations-statistics-key-facts-causes-and-insights/

5. LLMs as a radiologist ~ https://pubs.rsna.org/doi/abs/10.1148/radiol.240273
6. Advancing medical AI with Med-Gemini ~https://research.google/blog/advancing-medical-ai-with-med-gemini/