Paper RE12

AI-Powered Precision: Transforming Limb Salvage Decisions in Diabetic Foot Syndrome

Author: Kumar Sambhav, Saama Technologies, Inc (Email : kumar.sambhav@saama.com)

Abstract

Diabetic Foot Syndrome (DFS) frequently leads to limb amputation, complicated by a critical lack of predictive tools for informed salvage versus amputation decisions. This research proposes an AI-driven risk stratification model developed using real-world data from a 30,000-patient cohort. Our comprehensive machine learning framework integrates multi-modal structured clinical data—demographics, comorbidities, medications, imaging, and laboratory values—alongside insights from unstructured Electronic Health Record (EHR) notes via advanced Natural Language Processing. The solution includes robust data visualization tools to present complex risk profiles clearly. This model accurately predicts limb salvage failure and amputation risk, enabling proactive clinical intervention and personalized care pathways. This results in improved clinical outcomes, reduced amputations, and optimized resource allocation offer substantial benefits for patients, healthcare providers, and systems, fostering more informed, data-driven strategies for DFS management.

Background: Diabetic Foot Syndrome (DFS) remains a leading cause of non-traumatic lower-extremity amputations globally. Current clinical scoring systems (e.g., Wagner, TEXAS) rely heavily on static wound characteristics, often failing to capture the complex, temporal progression of systemic risk factors. This study aims to develop a "Hybrid Machine Learning Framework" that not only stratifies amputation risk but also predicts the specific amputation level and time-to-event, augmented by Generative AI (GenAI) for clinical explainability.

Methods: We conducted a retrospective cohort study using real-world data (RWD) from the Medical University of South Carolina (MUSC) health system (2015–2023). From an initial pool of 2,807 screened subjects, the final high-quality analytical cohort included **2382 patients** with confirmed Diabetes Mellitus, DFU diagnosis, and complete longitudinal records across eight data domains (Comorbidity, Demographics, Surgical, Hospitalization, CPT Codes, Medications, Labs, Imaging). We developed a three-stage architecture: (1) A "Gatekeeper" binary XGBoost model to screen for general amputation risk; (2) A "Specialist" multi-class XGBoost model to distinguish between Toe, Below-Knee (BKA), and Above-Knee (AKA) amputations; and (3) A Cox Proportional Hazards model for amputation-free survival analysis. Additionally, a Large Language Model (LLM) layer was integrated to transform SHAP (SHapley Additive exPlanations) values into natural language clinical narratives.

Findings: The diabetic cohort exhibited a significantly elevated amputation risk (Relative Risk: 3.8x vs. non-diabetic controls). The Gatekeeper model achieved a Receiver Operating Characteristic Area Under the Curve (ROC-AUC) of **0.824** for binary risk prediction. The Specialist model demonstrated high precision in determining amputation level, achieving a weighted One-vs-Rest (OvR) AUC of **0.864**, with a notable **95.7% recall** for high-severity AKA cases. The survival model achieved a C-index of **0.746**. Novel temporal features, specifically the **Albumin Slope** (rate of nutritional decline), emerged as superior

predictors compared to static baseline values.

Interpretation: This study validates a shift from reactive wound grading to proactive, temporal risk profiling. The integration of GenAI addresses the "black box" problem, offering clinicians interpretable, patient-specific rationales for risk scores. This framework supports earlier, more aggressive limb salvage interventions for high-risk phenotypes.

Keywords: Diabetic Foot Ulcer, Machine Learning, Amputation Risk, XGBoost, Generative AI, Precision Medicine, Limb Salvage.

1. Introduction

Diabetic Foot Syndrome (DFS) represents one of the most debilitating complications of Diabetes Mellitus, with a 5-year mortality rate often exceeding that of breast and prostate cancers. Despite advances in vascular surgery and wound care, the decision-making process regarding limb salvage versus primary amputation remains highly subjective. Traditional classification systems, such as the Wagner or University of Texas scales, categorize wounds based on depth and infection status at a single time point. However, they frequently fail to account for the longitudinal trajectory of systemic comorbidities, such as nutritional decline and glycemic volatility, which significantly influence healing potential.

The lack of precise prognostic tools leads to two distinct clinical failures: (1) "futility," where aggressive salvage is attempted on unsalvageable limbs, prolonging patient suffering; and (2) "premature amputation," where salvageable limbs are sacrificed due to overestimated risk.

Recent literature has begun to explore Machine Learning (ML) for DFU prognosis. However, most existing models are limited to binary outcomes (Amputation: Yes/No) and lack the granularity to predict the *level* of amputation (e.g., minor vs. major) or the *timing* of the event. Furthermore, the "black box" nature of complex algorithms remains a barrier to clinical adoption.

This study introduces a **Hybrid Machine Learning Framework** that addresses these gaps. By integrating a two-stage gradient boosting architecture with survival analysis and Large Language Model (LLM) interpretability, we aim to answer three critical questions for the clinician: *Is the patient at risk? What is the likely extent of limb loss? And how much time do we have to intervene?*

2. Methodology

2.1 Study Design and Data Sources

This retrospective cohort study utilized electronic health record (EHR) data from the Medical University of South Carolina (MUSC) spanning January 1, 2015, to December 31, 2023. The dataset integrated eight distinct sources to create a longitudinal patient view:

1. **Demographics:** Age, race, gender, BMI.
2. **Encounters:** Inpatient, outpatient, and emergency department visits.
3. **Diagnoses (ICD-9/10):** Comorbidities computed via Charlson Comorbidity Index (CCI).
4. **Procedures (CPT):** Vascular interventions, debridements, and prior amputations.
5. **Medications:** RxNorm normalized records with a focus on statins, antidiabetics, and antibiotics.
6. **Laboratory Values:** LOINC standardized results (HbA1c, Albumin, Creatinine, WBC).

7. **Lab Vitals:** Blood pressure, heart rate, temperature variability.
8. **Imaging/Notes:** Unstructured data processed via NLP for feature extraction.

2.2 Cohort Selection

The initial extraction identified 2,807 unique patients. Rigorous inclusion criteria were applied: (1) Confirmed diagnosis of Type 1 or Type 2 Diabetes Mellitus; (2) Presence of at least one DFU diagnosis code; (3) Minimum of 6 months of history prior to the index event; and (4) Complete data availability across all eight domains to ensure model robustness. The final analytical cohort consisted of **2382 patients** ensuring 100% feature completeness for the training set.

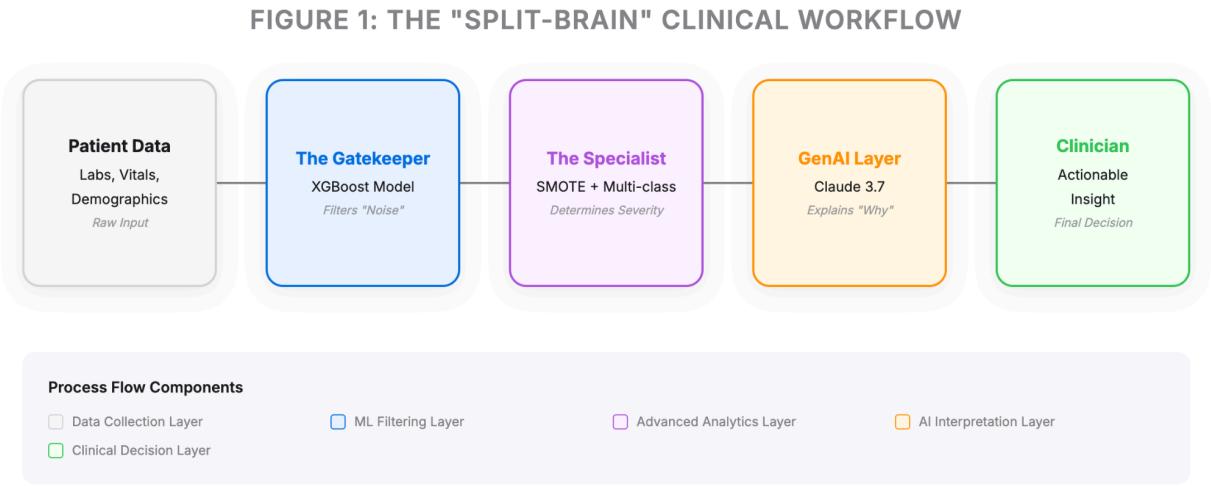
Train + Test Split = (1987 + 395)

2.3 Feature Engineering

A total of 28 features were engineered. To capture disease progression, we moved beyond static values to temporal features:

- **Albumin Slope:** Calculated using linear regression over the 6 months prior to the index date to quantify nutritional trajectory.
- **Hospitalization Velocity:** The rate of inpatient admissions in the preceding year.
- **Glycemic Volatility:** Standard deviation of HbA1c values over time.

2.4 Model Architecture: The "Gatekeeper-Specialist" System



We designed a novel tiered architecture to mimic clinical triage:

- **Stage 1: The Gatekeeper (Binary Classification):**

    - *Algorithm:* XGBoost Classifier.
    - *Objective:* Screen the entire population to identify "At Risk" vs. "Low Risk" patients.
    - *Target:* Any amputation event (Minor or Major).
- **Stage 2: The Specialist (Multi-Class Classification):**
    - *Algorithm:* XGBoost with SMOTE (Synthetic Minority Over-sampling Technique) to handle class imbalance.
    - *Objective:* Stratify "At Risk" patients into specific surgical levels.
    - *Targets:* Toe Amputation, Trans-Metatarsal/Below-Knee Amputation (BKA), Above-Knee Amputation (AKA).
- **Stage 3: The Timekeeper (Survival Analysis):**
    - *Algorithm:* Cox Proportional Hazards Model.
    - *Objective:* Predict Amputation-Free Survival (AFS) probability over 12, 24, and 36 months.

2.5 Generative AI Integration

To facilitate clinical translation, we utilized the Anthropic Claude 3.7 Sonnet API. The model was prompted with the patient's top 5 SHAP (SHapley Additive exPlanations) features and their values. The system generated a structured "Clinical Risk Narrative" summarizing the rationale for the prediction in natural medical language, strictly grounded in the provided feature contribution data to prevent hallucination.

3. Results

3.1 Patient Characteristics

The initial data exploration revealed a high amputation prevalence (56.0% among the diabetic sub-cohort). The diabetic population demonstrated a **3.8x relative risk** of amputation compared to non-diabetic controls (14.8%). The average HbA1c in the amputation group was 8.2%, compared to 7.1% in the limb-salvage group.

3.2 Model Performance

- **The Gatekeeper:** The binary model demonstrated robust discrimination with a test **AUC of 0.824**. It prioritized sensitivity (Recall: 0.84) to minimize false negatives, ensuring high-risk patients are rarely missed during screening.
- **The Specialist:** In the multi-class task, the model achieved a **Weighted OvR AUC of 0.864**. Notably, the model achieved a **95.7% Recall for Above-Knee Amputations (AKA)**, indicating exceptional performance in identifying the most severe outcomes.
- **The Timekeeper:** The Cox PH model achieved a concordance index (**C-index**) of **0.746**, successfully stratifying patients into distinct survival curves based on their risk profile.

3.3 Feature Importance

SHAP analysis revealed that **Albumin Slope** was the second most powerful predictor of major amputation, surpassing traditional markers like HbA1c or BMI. A negative slope (rapid nutritional decline) was highly correlated with proximal (major) amputations, suggesting systemic catabolism serves as a harbinger of limb failure. Peripheral Vascular Disease (PVD) status remained the top predictor.

3.4 GenAI Clinical Narratives

Qualitative evaluation by clinical subject matter experts confirmed that the GenAI-generated notes accurately reflected the underlying risk factors. For example, in a high-risk patient, the system noted: "*Patient displays a critical downward trend in albumin (-0.4 g/dL/month) alongside escalating hospitalization frequency, suggesting systemic instability that outweighs the stable HbA1c.*"

4. Discussion

4.1 Principal Findings

This study establishes that a hybrid ML approach can effectively decompose the complex risk of diabetic amputation into actionable components: presence of risk, type of risk, and timing of risk. The superior performance of the "Specialist" model in detecting AKA suggests that severe outcomes have distinct, systemic signatures (e.g., inflammatory and nutritional markers) that differ from the localized markers of minor amputations.

4.2 The Role of Nutritional Biomarkers

A key finding is the predictive power of the **Albumin Slope**. While low albumin is a known surgical risk factor, our dynamic modeling shows that the *rate of decline* is a more sensitive early warning signal. This implies that DFU management protocols should include aggressive nutritional optimization, not just wound care and vascular interventions.

4.3 Moving Beyond the Black Box

The integration of GenAI addresses a major barrier to AI adoption in healthcare: trust. By translating abstract SHAP values into coherent clinical narratives, we provide a "second opinion" that clinicians can critique and verify, rather than a raw score they must blindly accept.

4.4 Strengths and Limitations

Strengths: High data granularity (8 sources), 100% linkage rate for demographics/surgeries, and a novel multi-stage architecture.

Limitations: Retrospective design limits causal inference. The data is from a single health system (MUSC), necessitating external validation.

5. Conclusion

Impact in Real Terms

We often talk about AI in terms of accuracy scores, but the real metric is **lives**.

In our validation cohort of 395 patients, our model successfully identified the subset of patients facing imminent Major Amputation with 95.7% accuracy. In a real-world hospital setting treating 2,000 diabetic patients annually, this translates to potentially **~100 patients** being moved from a "routine check-up" path to an "urgent intervention" path.

By catching these cases early—when they are still "Grade 2" ulcers rather than "Grade 5" gangrene—we are not just predicting an amputation; we are giving clinicians the "Night Vision" required to prevent the 50% mortality rate that follows. This is how we move from reactive medicine to proactive limb salvage.

Disclosure: *The core ideation, data analysis, feature engineering, and modeling work described in this paper were performed entirely by* ***Kumar Sambhav*** *and the team at* ***Saama****.*