Original Paper

Admission-Time Prediction of In-Hospital Major Amputation in Diabetes-Related Foot Disease: Model Development and Temporal Validation Study

Yu-Tung Fu[1], MSc, MD; Yu-Yao Huang[2,3,4], MD; Jiun-Ting Yeh[1,3], MD; Ren-Wen Huang[1,3], MBA, MD

[1]Department of Plastic and Reconstructive Surgery, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[2]Division of Endocrinology and Metabolism, Department of Internal Medicine, Chang Gung Memorial Hospital, Taoyuan, Taiwan

[3]College of Medicine, Chang Gung University, Taoyuan, Taiwan

[4]Department of Medical Nutrition Therapy, Chang Gung Memorial Hospital, Taoyuan, Taiwan

Corresponding Author:
Ren-Wen Huang, MBA, MD
Department of Plastic and Reconstructive Surgery
Chang Gung Memorial Hospital
No. 5, Fuxing Street, Guishan District
Taoyuan, 333423
Taiwan
Phone: 886 3 3281200
Email: aaronhuang0327@gmail.com

Abstract

Background: Major above-ankle amputation is a devastating outcome of diabetes-related foot disease. The Wagner classification is routinely available but places many inpatients in the same grade; few prediction models for major amputation have undergone temporal validation or reported calibration.

Objective: This study aimed to develop an admission-time model for in-hospital major amputation from routinely collected data and to validate it internally in a later period at the same center. Its intended role is to support prognostic communication and prioritization of early multidisciplinary and vascular assessment and not to inform the decision to amputate. Secondary objectives were to compare machine learning algorithms with penalized logistic regression and to assess the added value of segmental limb pressures.

Methods: In this retrospective cohort study at a tertiary center in Taiwan, adults with a single hospitalization for diabetes-related foot disease from 2015 through 2025 were divided by admission date into a development cohort (2015-2021) and a temporal test cohort (2022-2025). Twenty admission variables were modeled. Five algorithms were compared by nested cross-validation in the development cohort; the one with the lowest out-of-fold log loss became the primary model, with operating thresholds fixed before a single evaluation in the test cohort.

Results: The development cohort included 1400 patients with 126 (9.0%) major amputations and the test cohort 771 patients with 51 (6.6%). In the test cohort, the primary ridge logistic regression model had an area under the receiver operating characteristic curve (AUROC) of 0.882 (95% CI 0.841-0.917) vs 0.804 (95% CI 0.737-0.867) for Wagner grade alone. Absolute risk was overpredicted: mean predicted risk was 9.6% vs 6.6% observed, with calibration intercept –0.47, slope 1.33, and observed-to-expected ratio 0.69 (95% CI 0.53-0.86). At the prespecified screening threshold, 343 (44.5%) patients were flagged; sensitivity was 0.980 (95% CI 0.897-0.997), specificity 0.593 (95% CI 0.557-0.628), positive predictive value 14.6% (95% CI 11.2%-18.7%), and negative predictive value 0.998 (95% CI 0.987-1.000). Of 771 patients, no major amputation occurred among the 396 (51.4%) patients in the lowest risk band; of 54 patients, 22 (40.7%) patients in the highest band underwent major amputation. Random forest and gradient boosting reached higher AUROCs (0.915 and 0.912) in secondary comparisons. Segmental limb pressures did not clearly improve discrimination in complete cases (change in AUROC –0.004).

Conclusions: In an internal temporal validation at a single center, a transparent model based on routine admission data added discrimination for in-hospital major amputation beyond Wagner grade alone but overpredicted absolute risk. It estimates amputation under 1 institution's care pathway and was not compared with multidimensional classifications such as Wound, Ischemia, and foot Infection; whether acting on it improves outcomes is untested. Independent external validation, recalibration where needed,

XSL•FO
RenderX

and prospective evaluation are required before clinical use; the model must not be used to indicate amputation or withhold limb-salvage evaluation.

(JMIR Med Inform 2026;14:e110740) doi: 10.2196/110740

KEYWORDS

diabetic foot; diabetes-related foot disease; lower-extremity amputation; clinical prediction model; machine learning; logistic regression; temporal validation; calibration; risk stratification; electronic health records

Introduction

Diabetes-related foot disease affects approximately 1 in 5 people with diabetes during their lifetime and remains a leading cause of nontraumatic lower-extremity amputation worldwide [1-4]. Ulcer recurrence is common even with multidisciplinary care [3], and 5-year mortality after major amputation approaches that of many common cancers [5,6]. Patients with diabetes-related foot disease may fear major amputation more than death [7].

The decision for major amputation is rarely immediate. Under multidisciplinary care, most limbs are first managed with debridement, antibiotic treatment, revascularization when indicated, and attempted reconstruction, and above-ankle amputation often comes only after several such procedures [8-12]. For patients and families, this staged course brings repeated procedures and prolonged uncertainty about the final outcome. The Wagner classification, the most widely available tool at admission, remains prognostic in limb-threatening disease [13-15]. It is, however, observer dependent, and it compresses wound depth, infection, and gangrene into a single grade with no measure of perfusion, the axis that the Wound, Ischemia, and foot Infection (WIfI) classification grades separately and that current classification guidance recommends for limb-threat assessment [10,16,17]. A risk estimate available at admission, before the first debridement, would help prepare the patient and family for what may follow and help the team prioritize early multidisciplinary review.

Prediction models could provide such estimates. A meta-analysis of machine learning models for amputation in diabetes-related foot disease reported pooled discrimination of approximately 0.81 [18], but few such models [19,20] have undergone temporal validation, that is, evaluation unchanged in a later period [21-23], and calibration of absolute risk is seldom reported although it determines whether a predicted probability can be acted on [24,25]. Current guidance emphasizes prespecification, separation of development from evaluation, and reporting of calibration and clinical usefulness alongside discrimination [24,26-28]. Two further distinctions shape how such a model should be read. First, in a service in which every patient receives structured multidisciplinary care, major amputation is a care-dependent outcome: a model predicts the probability of amputation under the observed pathway, not an untreated biological risk of limb loss [29,30]. Second, accurate prediction does not by itself improve outcomes; whether acting on a prediction changes the timing of care or prevents amputation is a separate question that requires an impact study [31,32].

Our pathway assesses perfusion in every admitted patient, but medial arterial calcification can make cuff-based pressures falsely reassuring in diabetes [10,33,34], so whether routinely measured segmental limb pressures add prognostic information beyond clinical variables is uncertain.

We therefore aimed to develop an admission-time model for in-hospital major amputation from routinely collected data and to validate it internally in a later period at the same center, within a mature, protocolized multidisciplinary diabetes-related foot service. The intended clinical role, stated in advance, is to support prognostic communication with patients and families and the prioritization of early multidisciplinary and vascular assessment, not to inform the decision to amputate or to withhold limb-salvage care. Our hypothesis was that the primary model, chosen among 5 algorithms by a prespecified selection rule, would discriminate major amputation better than Wagner grade alone, the benchmark universally available at admission. Secondary objectives were to compare flexible machine learning algorithms with the primary model and to assess whether wound-side segmental limb pressures add prognostic information. An exploratory learning-curve analysis examined how the number of development cohort events affected the relative performance of logistic regression and gradient boosting.

Methods

Study Design and Setting

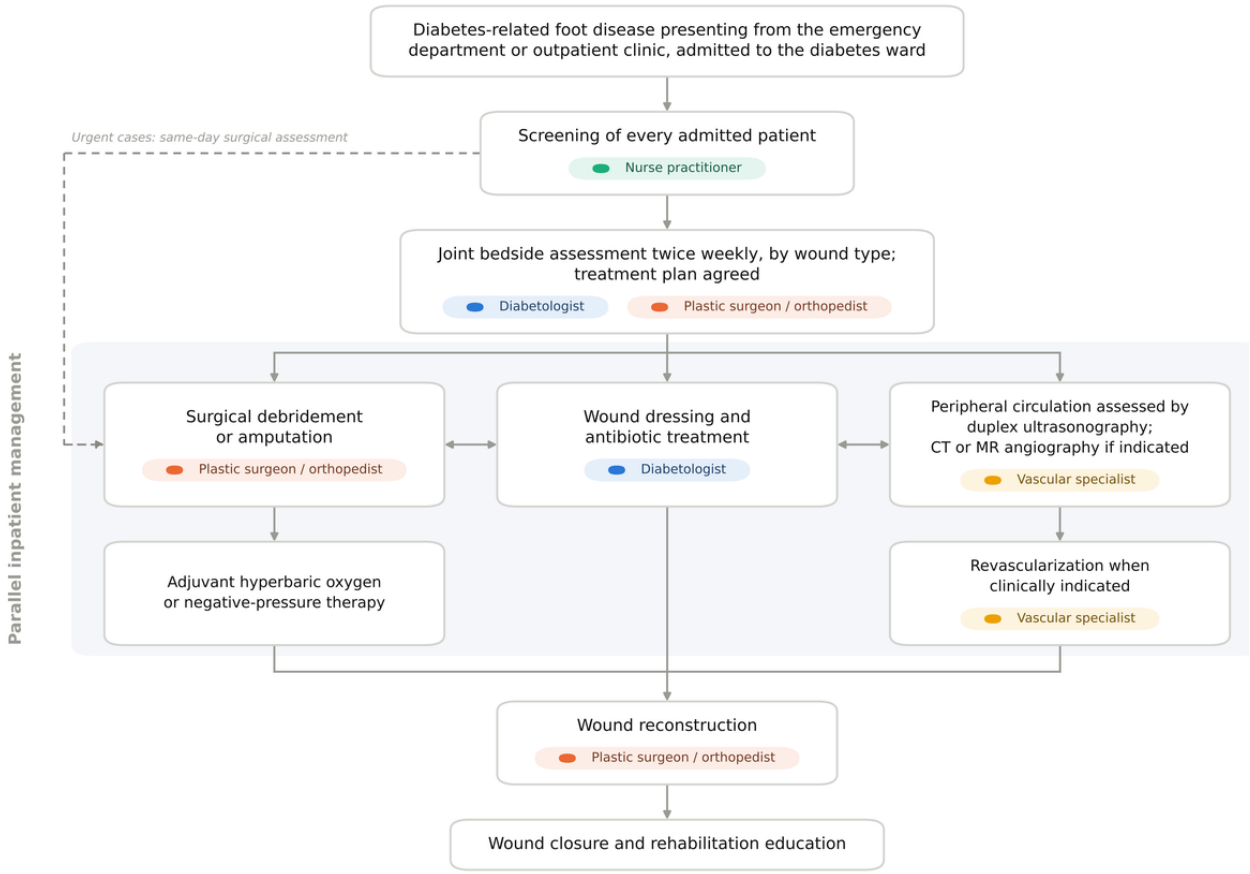
We conducted a retrospective cohort study at Chang Gung Memorial Hospital, a tertiary referral center in Taiwan, using electronic health record data from January 2015 through December 2025. Reporting follows the TRIPOD+AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + AI) statement [26]; the completed checklist is provided in Multimedia Appendix 1.

All patients admitted with diabetes-related foot disease entered a standardized multidisciplinary pathway that remained in place throughout the study period (Figure 1). A diabetes-related foot nurse practitioner screened every admission. Twice weekly, a diabetologist and either a plastic surgeon or an orthopedic surgeon, assigned according to wound type, performed a joint bedside assessment and agreed on the treatment plan; patients identified at screening as needing urgent attention were referred for same-day surgical assessment rather than waiting for the next scheduled round. Inpatient care proceeded in parallel: source control by debridement or amputation when required, wound care and antimicrobial treatment, and assessment of peripheral circulation, which the protocol specifies for every admitted patient, by duplex arterial ultrasonography with computed tomography or magnetic resonance angiography when

XSL•FO
RenderX

indicated. A vascular specialist performed revascularization when clinically indicated. Selected patients received negative-pressure wound therapy or hyperbaric oxygen therapy, followed, when feasible, by reconstruction, wound closure, and rehabilitation education. Podiatry is not established as a separate profession in Taiwan, so the assessment and wound care roles it holds in many North American and European foot services are performed here by the nurse practitioner, plastic surgeons, and orthopedic surgeons.

Figure 1. Institutional multidisciplinary care pathway for patients admitted with diabetes-related foot disease, in place throughout the study period. Colored tags identify the discipline responsible for each step. The dashed path indicates escalation of urgent cases identified at screening to same-day surgical assessment, without waiting for the next scheduled joint round. The protocol specifies peripheral circulation assessment by duplex ultrasonography for every admission; CT or MR angiography and revascularization are undertaken when clinically indicated. CT: computed tomography; MR: magnetic resonance.



Ethical Considerations

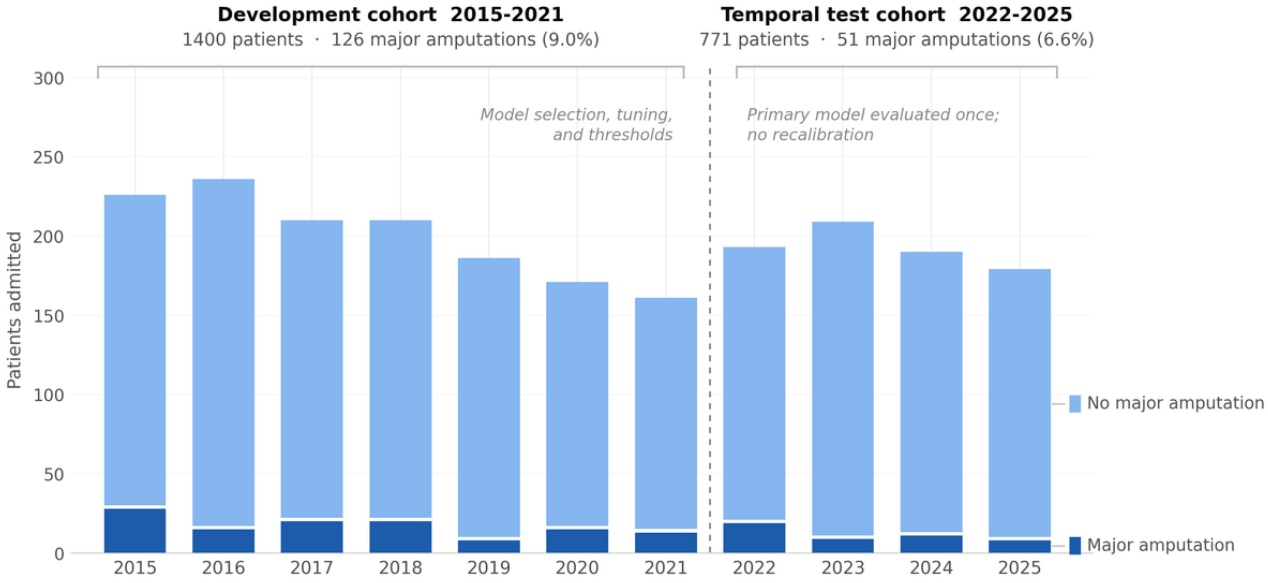
The Institutional Review Board of the Chang Gung Medical Foundation approved the study (number 202600100B0) and waived informed consent because of the retrospective design and use of deidentified data. All data were deidentified before analysis. The study was conducted in accordance with the Declaration of Helsinki. It was not prospectively registered; the predictor set, temporal split, primary-model selection rule, and operating thresholds were fixed before the test cohort was evaluated.

Participants and Temporal Split

Eligible participants were adults admitted to the diabetes-related foot service with diabetes-related foot disease who had a single such hospitalization during the study period; each patient therefore contributed 1 hospitalization and the unit of analysis was the patient (Figure 2). Because this criterion depends on the absence of later admissions, which is not known at the index admission, post hoc sensitivity analyses (1) repeated the development and evaluation procedure in the first eligible admission of every patient and (2) applied the primary model unchanged to those admissions. Admissions from 2015 through 2021 formed the development cohort, and admissions from 2022 through 2025 formed the temporal test cohort. No patient contributed to both cohorts. All preprocessing, model selection, hyperparameter tuning, and threshold selection were completed in the development cohort. The primary model and its thresholds were evaluated once in the test cohort, without model updating or recalibration; the test cohort was subsequently reused only for the secondary, sensitivity, and exploratory analyses, which are labeled as such. Patients with bilateral ulceration were retained as 1 patient-level observation with a binary bilateral-disease indicator. Because amputation laterality was unavailable in the source data, a sensitivity analysis excluded bilateral cases.

XSL•FO
RenderX

Figure 2. Cohort flow and temporal split. Of 2941 adults admitted with diabetes-related foot disease in 2015-2025 (4347 admissions), 770 patients with more than 1 hospitalization (2176 admissions) were excluded, leaving 2171 adults with a single hospitalization, of whom 177 (8.2%) underwent in-hospital major amputation. Columns show these 2171 patients by calendar year of admission, and darker segments indicate in-hospital major amputation. The 2015-2021 development cohort was used for preprocessing, model selection, hyperparameter tuning, and threshold derivation. The primary model was evaluated once in the 2022-2025 temporal test cohort, without recalibration.



Outcome and Prediction Time

The outcome was in-hospital major amputation, defined as transtibial (below-knee) or transfemoral (above-knee) amputation of either lower limb and ascertained from operative records; laterality was not recorded. All forefoot and midfoot procedures were recorded separately as minor amputation and were not counted as the outcome. The prediction time was admission. Predictors were restricted to information recorded at admission, before debridement, revascularization, or amputation. The intended use is to support early prognostic assessment, communication with the patient and family, and prioritization of multidisciplinary review. The model is not intended to determine candidacy for amputation or limb salvage. Because every patient in the cohort received structured multidisciplinary care, including vascular assessment and revascularization when indicated, the model estimates the probability of in-hospital major amputation under this care pathway rather than an untreated risk of limb loss.

Candidate Predictors

The prespecified model included 20 variables available at admission: age, sex, BMI, and ever-smoking; hypertension, ischemic heart disease, congestive heart failure, prior stroke, chronic kidney disease, and osteomyelitis; hemoglobin, white blood cell count, neutrophil percentage, C-reactive protein, albumin, blood urea nitrogen, creatinine, and glycated hemoglobin; Wagner grade; and bilateral involvement. White blood cell count, C-reactive protein, and creatinine were log-transformed. Sex was obtained from administrative records and refers to sex assigned at birth; gender identity was not captured.

For the prespecified secondary analysis, we added wound-side segmental limb pressures at 4 cuff levels: high thigh, low thigh, calf, and ankle. These were measured with pneumatic cuffs and continuous-wave Doppler during the lower-extremity arterial physiologic study, of which duplex scanning is the imaging component. For bilateral disease, the lower value from the 2 limbs was used at each level. Toe pressure, toe-brachial index, and transcutaneous oxygen pressure, the perfusion measures least affected by medial arterial calcification, were not recorded. Brachial pressures were recorded, so an ankle-brachial ratio was derivable in 1503 of 2171 (69.2%) patients; it was not among the prespecified predictors and, being cuff-based, is subject to the same calcification artifact as the segmental pressures.

Missing Data

Implausible values (ages outside 18-105 years and laboratory values outside predefined plausibility limits) were set to missing before imputation. Missing predictor values were imputed by multivariate iterative imputation fitted in the development cohort and then applied unchanged to the test cohort. Standardization also used development cohort parameters. Missingness was summarized by cohort and outcome (Multimedia Appendix 2). No patient was excluded from the primary analysis because of a missing predictor. In particular, restricting the cohort to patients with complete wound-side pressure measurements would have removed 48.0% (85/177) of all outcome events and selected a subgroup whose amputation rate was substantially lower.

For prospective application to an individual patient, the imputation model, the standardization parameters, and the ridge coefficients fitted in the development cohort would be applied unchanged: a missing value for any of the 20 predictors would be imputed from the patient's observed predictors by the development-fitted imputer, and none of these components would be refitted using the new patient's data or any information arising after the prediction time. Albumin and glycated hemoglobin were the predictors most often unavailable, and

XSL•FO
RenderX

glycated hemoglobin missingness increased from 7.6% to 21.9% between eras; the post hoc sensitivity analyses that refitted the model without each of these predictors (see Sensitivity Analyses in the Results section) therefore describe how far the model depends on them. Rules for withholding an output when inputs are extensively missing or implausible, and monitoring of missingness patterns over time, would need to be specified and tested during prospective evaluation; clinical assessment should not depend on the availability of a score.

Sample Size

We used all eligible patients in the study window rather than selecting a target sample size. We assessed data adequacy post hoc using the criteria of Riley et al [35,36], which were developed for regression models and were applied to the logistic regression model. With 20 predictor parameters, an outcome proportion of 9.0%, and an apparent Cox-Snell R^2 of 0.175 for the unpenalized model, the estimated minimum sample size to limit expected shrinkage to 10% was 925 patients, compared with 1400 available; the criteria for a small difference between apparent and adjusted Nagelkerke R^2 (0.05) and for precise estimation of the overall risk (margin 0.05) required 791 and 126 patients, respectively. The development cohort provided 6.3 events per candidate predictor. Because the calculation used an apparent R^2 estimated from the same data, it is a plausibility check rather than a prospective justification and does not establish adequacy for the more data-hungry flexible algorithms.

Model Development and Prespecified Selection

We compared ridge-penalized logistic regression, random forest, gradient boosting, histogram-based gradient boosting, and a multilayer perceptron neural network. Hyperparameters were selected by 5-fold cross-validation in the development cohort, optimizing log loss, with imputation and standardization embedded in the modeling pipeline and refitted within every training fold, so that no preprocessing information crossed fold boundaries; search grids are reported in Multimedia Appendix 3. Models were fitted at the observed event prevalence without resampling, class weighting, or post hoc probability correction [37]. Two clinical benchmarks were fitted within the same framework: logistic regression with Wagner grade entered as a categorical factor, and logistic regression with Wagner grade entered as a single linear predictor.

The primary algorithm was selected before any test–cohort evaluation as the model with the lowest out-of-fold log loss under nested cross-validation (5 outer evaluation folds and an inner 5-fold hyperparameter search within each outer training fold). Two operating thresholds were fixed from the primary model's out-of-fold predictions: the highest threshold retaining at least 90% sensitivity, intended for screening, and the Youden-maximizing threshold. Because positive predictive value depends on prevalence, we also prespecified thresholds corresponding to the highest 10% and 5% of out-of-fold predicted risk and risk bands of <5%, 5% to 15%, 15% to 30%, and ≥30%. The other 4 algorithms were evaluated as secondary comparisons.

Exploratory Learning Curve

We explored whether development cohort size altered the relative performance of logistic regression and gradient boosting. Random subsamples containing 20%, 35%, 50%, 65%, or 80% of the development cohort were drawn 100 times at each fraction and evaluated on the unchanged temporal test cohort. Both algorithms were retuned by 5-fold cross-validation within every subsample; in this exploratory analysis, imputation and standardization were fitted once per subsample and applied identically to both algorithms. For each replicate, we recorded the paired difference in test area under the receiver operating characteristic curve (AUROC); the 2.5th and 97.5th percentiles describe subsampling variability. Because all replicates shared the same test cohort with 51 events, this analysis describes this dataset and does not establish a general event-count threshold.

Statistical Analysis

Baseline characteristics were summarized by cohort. Case-mix differences were quantified with standardized mean differences (SMDs), calculated for continuous variables as the absolute difference in means divided by the square root of the mean of the 2 cohort variances, with Wagner grade treated as an ordinal numeric score (0-5), and for binary variables as the absolute difference in proportions divided by the square root of the mean of the 2 binomial variances; for presenting wound type, each category was treated as a binary indicator and given its own SMD. Available cases were used for each variable. Discrimination was assessed with the AUROC and area under the precision-recall curve (AUPRC), overall accuracy with the Brier score and scaled Brier score, and calibration with the intercept, slope, observed-to-expected ratio, and a grouped plot by decile of predicted risk. No recalibration was performed in the test cohort. CIs for AUROC, AUPRC, calibration slope, and observed-to-expected ratio were obtained from 2000 patient-level bootstrap resamples. Paired model comparisons used differences calculated on identical resamples. Wilson score intervals were used for proportions, including sensitivity, specificity, and positive and negative predictive values at each operating threshold, computed from the confusion-matrix counts. Permutation importance was calculated in the test cohort as the mean decrease in AUROC across 50 permutations of each variable, reported with its standard deviation.

Clinical usefulness was evaluated by decision curve analysis [38-40], using Wagner grade alone, an assess-all strategy, and an assess-none strategy as comparators. A positive prediction was defined to trigger escalated early assessment, including urgent multidisciplinary and vascular review, rather than an amputation decision. Threshold probabilities from 1% to 40% were examined. Because the proposed action is an earlier assessment within a pathway that already assesses every patient, rather than an invasive treatment, the harm of a false-positive prediction is modest and the clinically relevant thresholds are low: the lower part of the range (approximately 1%-10%) corresponds to a screening role in which clinicians accept escalating assessment in many patients to avoid missing an amputation, and it contains the prespecified screening (5.9%) and Youden (8.6%) thresholds. Higher thresholds correspond to increasingly selective prioritization and were included to

XSL•FO
RenderX

characterize the model's behavior up to the highest prespecified risk band (≥30%); they are not proposed as clinical decision thresholds. The range was chosen by the investigators and was not derived from elicited clinician or patient preferences. For the secondary vascular analysis, the incremental value of the 4 segmental limb pressures was estimated in patients with complete wound-side measurements in both cohorts. We also described potential mechanisms of vascular-data missingness by comparing patients with and without recorded pressures.

Prespecified sensitivity analyses refitted the primary algorithm after excluding in-hospital deaths, bilateral cases, and admissions in 2020-2021, and for a composite outcome of major amputation or in-hospital death. Six further sensitivity analyses were not prespecified and were performed post hoc: refitting without glycated hemoglobin and refitting without albumin to test dependence on the 2 predictors whose measurement properties differ most between eras or between patients; refitting and evaluating only in patients with an observed albumin value, because albumin was ordered preferentially in sicker patients; refitting with the recorded presenting wound type added to the 20 prespecified variables; repeating the development and temporal evaluation in the first eligible admission of every patient, including patients with more than 1 hospitalization; and applying the primary model unchanged to the first eligible admission of every test-period patient. Analyses used Python (version 3.11; Python Software Foundation) with scikit-learn (version 1.8.0; scikit-learn developers) and statsmodels (version 0.15.0; statsmodels developers). Random seeds were fixed. Analysis code and model artifacts are available as described in the Data Availability section.

Results

Cohort

Of 2941 patients admitted with diabetes-related foot disease during the study period (4347 admissions), 770 (26.2%) had more than 1 hospitalization (2176 admissions) and were excluded. The study therefore included 2171 patients, corresponding to 49.9% of all admissions: 1400 patients with 126 (9.0%) major amputations in the development cohort and 771 patients with 51 (6.6%) major amputations in the temporal test cohort (Figure 2 and Table 1). Of the 177 major amputations, 112 were transtibial and 65 transfemoral; minor amputation was performed in 324 of 1400 (23.1%) development cohort patients and 221 of 771 (28.7%) test cohort patients. Compared with the development cohort, the test cohort had a lower prevalence of chronic kidney disease (23.2% vs 30.9%; SMD 0.17), more osteomyelitis (23.7% vs 16.4%; SMD 0.19), and a different Wagner distribution. Grade 3 accounted for 43.8% and 55.3% of the development and test cohorts, respectively; grade 5 for 3.4% and 6.6%; and necrotizing fasciitis at presentation for 8.1% and 17.3%. Median length of stay did not shorten between eras (22 vs 25 days). Annual event rates are reported in Multimedia Appendix 4.

XSL•FO
RenderX

Table 1. Baseline characteristics of the development and temporal test cohorts[a].

Characteristic	Development 2015-2021 (n=1400)	Missing, n (%)	Temporal test 2022-2025 (n=771)	Missing, n (%)	SMD[b]
Age, years	66.5 (57.1-75.7)	3 (0.2)	65.8 (55.9-75.3)	3 (0.4)	0.03
Male sex	849 (60.6)	0 (0.0)	514 (66.7)	0 (0.0)	0.13
Ever smoker	297 (21.2)	2 (0.1)	181 (23.5)	0 (0.0)	0.05
BMI, kg/m²	24.4 (21.9-27.8)	33 (2.4)	25.0 (22.5-28.7)	9 (1.2)	0.12
Comorbidity (chart diagnosis)					
Hypertension	990 (70.7)	0 (0.0)	521 (67.6)	0 (0.0)	0.07
Ischemic heart disease	347 (24.8)	0 (0.0)	183 (23.7)	0 (0.0)	0.02
Congestive heart failure	102 (7.3)	0 (0.0)	66 (8.6)	0 (0.0)	0.05
Prior stroke	214 (15.3)	0 (0.0)	96 (12.5)	0 (0.0)	0.08
Chronic kidney disease	432 (30.9)	0 (0.0)	179 (23.2)	0 (0.0)	0.17
Renal failure	302 (21.6)	0 (0.0)	193 (25)	0 (0.0)	0.08
Osteomyelitis	229 (16.4)	0 (0.0)	183 (23.7)	0 (0.0)	0.19
Wound at admission					
Wagner grade, median (IQR)	3.0 (2.0-4.0)	0 (0.0)	3.0 (3.0-4.0)	0 (0.0)	0.28
Grade 0	45 (3.2)	NA[c]	48 (6.2)	NA	NA
Grade 1	214 (15.3)	NA	13 (1.7)	NA	NA
Grade 2	111 (7.9)	NA	17 (2.2)	NA	NA
Grade 3	613 (43.8)	NA	426 (55.3)	NA	NA
Grade 4	369 (26.4)	NA	216 (28)	NA	NA
Grade 5	48 (3.4)	NA	51 (6.6)	NA	NA
Presenting wound type	NA	1 (0.1)	NA	0 (0.0)	NA
Cellulitis	445 (31.8)	NA	321 (41.6)	NA	0.21
Subcutaneous tissue	129 (9.2)	NA	6 (0.8)	NA	0.39
Tendon	10 (0.7)	NA	0 (0.0)	NA	0.12
Bone	404 (28.9)	NA	157 (20.4)	NA	0.20
Gangrene	297 (21.2)	NA	154 (20)	NA	0.03
Necrotizing fasciitis	114 (8.1)	NA	133 (17.3)	NA	0.28
Bilateral ulcers	174 (12.4)	0 (0.0)	108 (14)	0 (0.0)	0.05
Admission laboratory values[d]					
Hemoglobin, g/dL	11.0 (9.5-12.7)	5 (0.4)	11.0 (9.3-12.8)	6 (0.8)	0.02
White blood cell count, $\times 10^3/\mu L$	10.7 (8.0-15.3)	5 (0.4)	11.0 (7.9-15.4)	8 (1.0)	0.02
Neutrophils, %	77.8 (70.0-85.0)	8 (0.6)	80.0 (72.3-86.0)	5 (0.6)	0.17
CRP[e], mg/L	55.9 (13.7-143.9)	21 (1.5)	73.1 (14.4-160.7)	7 (0.9)	0.12
Albumin, g/dL	3.2 (2.8-3.6)	395 (28.2)	3.3 (2.8-3.6)	178 (23.1)	0.11
BUN[f], mg/dL	23.5 (15.4-40.9)	28 (2)	26.1 (17.0-43.0)	11 (1.4)	0.07
Creatinine, mg/dL	1.3 (0.9-3.2)	14 (1)	1.3 (0.9-3.2)	11 (1.4)	0.01
HbA_{1c}[g], %	8.2 (6.9-10.3)	106 (7.6)	8.0 (6.9-10.3)	169 (21.9)	0.04
In-hospital outcomes					
Major amputation	126 (9)	0 (0.0)	51 (6.6)	0 (0.0)	0.09
Transtibial	77 (5.5)	NA	35 (4.5)	NA	NA

XSL•FO
RenderX

Characteristic	Development 2015-2021 (n=1400)	Missing, n (%)	Temporal test 2022-2025 (n=771)	Missing, n (%)	SMD[b]
Transfemoral	49 (3.5)	NA	16 (2.1)	NA	NA
Minor amputation	324 (23.1)	0 (0.0)	221 (28.7)	0 (0.0)	0.13
Death	98 (7)	0 (0.0)	46 (6)	0 (0.0)	0.04
Length of stay, days	22 (13-39)	0 (0.0)	25 (14-42)	0 (0.0)	0.08

[a]Values are median (IQR) or n (%). Comorbidities were recorded as present or absent from the chart diagnosis, so absence of a diagnosis was coded as absent. Chronic kidney disease is the documented comorbidity and was a prespecified predictor; renal failure denotes diabetic nephropathy recorded at the renal failure stage (end-stage kidney disease or dialysis), the most advanced of the 4 recorded stages (none, incipient proteinuria, overt proteinuria, and renal failure), and is descriptive only. Presenting wound type is a single mutually exclusive descriptor recorded separately from Wagner grade; it was not among the 20 prespecified predictors.

[b]SMD: standardized mean difference.

[c]NA: not applicable.

[d]SI conversions: hemoglobin and albumin (g/dL) × 10=g/L; BUN (mg/dL) × 0.357=mmol/L; creatinine (mg/dL) × 88.4=µmol/L; white blood cell count ($\times 10^3$/µL) = $\times 10^9$/L; HbA_{1c} (%): [HbA_{1c} – 2.15] × 10.929=mmol/mol.

[e]CRP: C-reactive protein.

[f]BUN: blood urea nitrogen.

[g]HbA_{1c}: hemoglobin A_{1c}.

Primary Model Selection in the Development Cohort

In nested cross-validation within the development cohort, ridge logistic regression had the lowest out-of-fold log loss (0.219; AUROC 0.869), followed by the neural network (0.223; 0.864), random forest (0.227; 0.857), gradient boosting (0.229; 0.851), and histogram-based gradient boosting (0.240; 0.850). Ridge logistic regression was therefore designated the primary model. Its out-of-fold predictions defined the screening (0.059), Youden (0.086), top 10% (0.264), and top 5% (0.371) thresholds before the temporal test cohort was evaluated.

Performance in the Temporal Test Cohort

In the temporal test cohort, the primary model achieved an AUROC of 0.882 (95% CI 0.841-0.917), an AUPRC of 0.392 (95% CI 0.271-0.532; no-skill value 0.066), a Brier score of 0.0512, and a scaled Brier score of 0.17. The model overpredicted risk: the calibration intercept was –0.47, the slope was 1.33 (95% CI 1.07-1.67), the observed-to-expected ratio was 0.69 (95% CI 0.53-0.86), and mean predicted risk was 9.6% compared with an observed event rate of 6.6%. All 5 multivariable algorithms discriminated better than the Wagner-grade benchmark, which achieved an AUROC of 0.804 (95% CI 0.737-0.867) whether the grade was entered as a categorical factor or as a single linear predictor; the primary model exceeded the categorical benchmark by 0.078 (95% CI 0.012-0.142). Discrimination is shown in Table 2 and Figure 3 (receiver operating characteristic curves in Figure 3A and precision-recall curves in Figure 3B); calibration and paired comparisons with the primary model are shown in Table 3 and the calibration plot in Figure 4A.

XSL•FO
RenderX

Table 2. Discrimination and overall accuracy of the models in the 2022-2025 temporal test cohort (771 patients, 51 events), with the development out-of-fold log loss used for model selection. All models were developed in the 2015-2021 cohort (1400 patients, 126 events) and evaluated without recalibration[a,b].

Model	Development OOF[c] log loss	AUROC[d] (95% CI)	AUPRC[e] (95% CI)	Brier score
Discrimination and overall accuracy				
Logistic regression (primary)	0.219	0.882 (0.841-0.917)	0.392 (0.271-0.532)	0.0512
Neural network	0.223	0.870 (0.826-0.910)	0.358 (0.243-0.501)	0.0532
Random forest	0.227	0.915 (0.878-0.946)	0.433 (0.314-0.597)	0.0481
Gradient boosting	0.229	0.912 (0.877-0.942)	0.506 (0.369-0.644)	0.0468
Histogram-based gradient boosting	0.240	0.903 (0.867-0.933)	0.414 (0.293-0.550)	0.0509
Wagner grade, categorical (benchmark)	NA[f]	0.804 (0.737-0.867)	0.314 (0.205-0.441)	0.0506
Wagner grade, linear (benchmark)	NA	0.804 (0.737-0.867)	0.314 (0.205-0.441)	0.0526

[a]Models are ordered by development OOF log loss, the prespecified selection criterion; the benchmarks follow.

[b]CIs were obtained from 2000 patient-level bootstrap resamples; 1 common set of resamples was used for all AUROC intervals and paired differences, and a second common set for all AUPRC intervals. The 2 benchmark specifications rank patients identically, so their discrimination is identical; they differ only in calibration (Table 3).

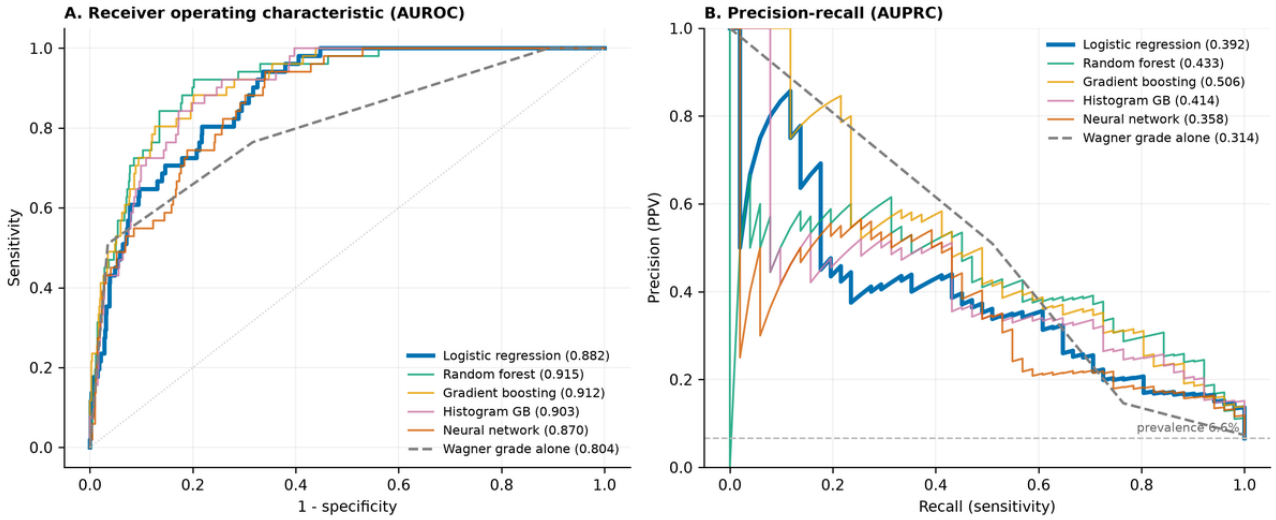
[c]OOF: out-of-fold.

[d]AUROC: area under the receiver operating characteristic curve.

[e]AUPRC: area under the precision-recall curve (no-skill value = prevalence, 0.066).

[f]Not applicable.

Figure 3. Discrimination in the temporal test cohort (2022-2025; 771 patients, 51 events). (A) Receiver operating characteristic curves for the 5 multivariable algorithms and the Wagner-only benchmark; the prespecified primary logistic regression model is emphasized. (B) Precision-recall curves; the horizontal dashed line indicates the event prevalence of 6.6%. AUPRC: area under the precision-recall curve; AUROC: area under the receiver operating characteristic curve; GB: gradient boosting; PPV: positive predictive value.



XSL•FO
RenderX

Table 3. Calibration of the models and paired comparison of discrimination with the primary model in the 2022-2025 temporal test cohort (771 patients, 51 events), without recalibration[a,b].

Model	Calibration slope	O:E[c] ratio	Paired ΔAUROC[d] vs primary (95% CI)
Calibration and paired comparison with the primary model			
Logistic regression (primary)	1.33	0.69	NA[e]
Neural network	0.91	0.71	–0.012 (–0.039 to 0.014)
Random forest	1.68	0.66	+0.033 (0.009 to 0.062)
Gradient boosting	1.22	0.65	+0.030 (0.005 to 0.059)
Histogram-based gradient boosting	0.98	0.70	+0.021 (–0.007 to 0.049)
Wagner grade, categorical (benchmark)	1.44	0.56	–0.078 (–0.142 to –0.012)
Wagner grade, linear (benchmark)	1.60	0.60	–0.078 (–0.142 to –0.012)

[a]Models are ordered by development out-of-fold log loss, the prespecified selection criterion; the benchmarks follow.

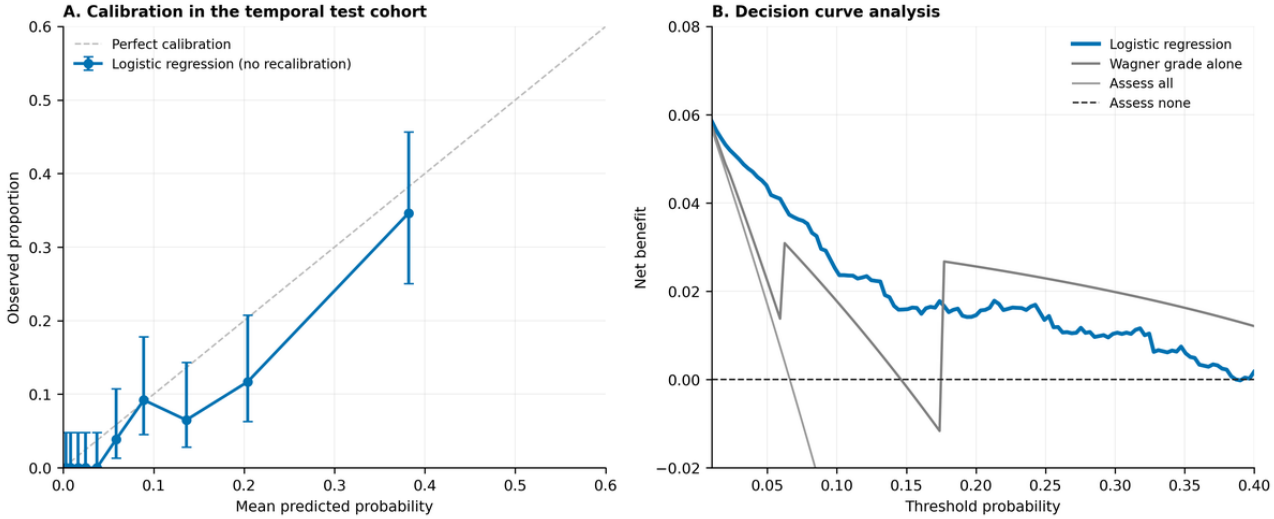
[b]CIs for the paired differences were obtained from the same 2000 patient-level bootstrap resamples as in Table 2. The 2 benchmark specifications rank patients identically, so their paired comparisons with the primary model are identical; they differ only in calibration.

[c]O:E: observed-to-expected ratio.

[d]ΔAUROC: paired difference in the area under the receiver operating characteristic curve (AUROC) between each model and the primary model.

[e]Not applicable.

Figure 4. Calibration and clinical usefulness of the primary model in the temporal test cohort. (A) Grouped calibration plot by decile of predicted risk, without recalibration. Error bars show 95% Wilson intervals; the diagonal dashed line indicates perfect calibration. The numbers of patients and major amputation events in each decile are reported in Table S6 in Multimedia Appendix 4. (B) Decision curve analysis comparing the primary logistic regression model, Wagner grade alone, assess-all, and assess-none strategies across threshold probabilities of 1% to 40%. A positive prediction represents escalation to earlier multidisciplinary assessment, not a decision to amputate; thresholds below approximately 10% correspond to the screening role proposed for the model, and higher thresholds describe increasingly selective prioritization.



In secondary test-cohort comparisons, random forest achieved an AUROC of 0.915 (95% CI 0.878-0.946) and gradient boosting an AUROC of 0.912 (95% CI 0.877-0.942). The paired differences from the primary model were +0.033 (95% CI 0.009-0.062) for random forest and +0.030 (95% CI 0.005-0.059) for gradient boosting; the differences did not clearly exclude zero for histogram-based gradient boosting (+0.021, 95% CI –0.007 to 0.049) or the neural network (–0.012, 95% CI –0.039 to 0.014). Because these algorithms were compared in the test cohort, the findings are descriptive secondary observations and were not used to redesignate the primary model.

Risk Stratification and Predictive Values

At the prespecified screening threshold, sensitivity was 0.980 (95% CI 0.897-0.997), specificity was 0.593 (95% CI 0.557-0.628), positive predictive value was 14.6% (95% CI 11.2%-18.7%), and negative predictive value was 0.998 (95% CI 0.987-1.000). Of 771 patients, the threshold flagged 343 (44.5%) patients; 50 of 51 major amputations occurred above it, and 293 flagged patients did not undergo major amputation. At the Youden threshold, sensitivity was 0.902 (95% CI 0.790-0.957) and specificity was 0.675 (95% CI 0.640-0.708). Of 771 patients, the top 10% threshold flagged 75 (9.7%) patients with a positive predictive value of 34.7% (95% CI 24.9%-45.9%); of 771 patients, the top 5% threshold flagged

XSL•FO
RenderX

36 (4.7%) patients with a positive predictive value of 41.7% (95% CI 27.1%-57.8%; Table 4).

Table 4. Prespecified operating thresholds for the primary logistic regression model in the 2022-2025 temporal test cohort (771 patients, 51 events)[a].

Operating point	Threshold	Flagged, n (%)	Sensitivity (95% CI)	Specificity (95% CI)	PPV[b], % (95% CI)	NPV[c] (95% CI)
Prespecified operating thresholds						
Screening (OOF[d] sensitivity ≥90%)	0.059	343 (44.5)	0.980 (0.897-0.997)	0.593 (0.557-0.628)	14.6 (11.2-18.7)	0.998 (0.987-1.000)
Youden index maximum	0.086	280 (36.3)	0.902 (0.790-0.957)	0.675 (0.640-0.708)	16.4 (12.5-21.2)	0.990 (0.976-0.996)
Top 10% of OOF risk	0.264	75 (9.7)	0.510 (0.377-0.641)	0.932 (0.911-0.948)	34.7 (24.9-45.9)	0.964 (0.948-0.976)
Top 5% of OOF risk	0.371	36 (4.7)	0.294 (0.187-0.430)	0.971 (0.956-0.981)	41.7 (27.1-57.8)	0.951 (0.933-0.964)

[a]Thresholds were fixed in the development cohort from OOF predictions of the primary model. All CIs are Wilson 95% CIs computed from the confusion-matrix counts at each threshold (51 events and 720 nonevents); for example, at the screening threshold there were 50 true positives, 293 false positives, 427 true negatives, and 1 false negative.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

[d]OOF: out-of-fold.

Observed event rates increased across the prespecified risk bands (Table 5). Of 771 patients, no major amputation occurred among the 396 (51.4%) patients with predicted risk <5% (0.0%, 95% CI 0.0%-1.0%). In the ≥30% band, of 54 patients, 22 (40.7%, 95% CI 28.7%-54.0%) underwent major amputation and 32 (59.3%) did not. Decision curve analysis (Figure 4B) showed greater net benefit for the model than for assess-all and assess-none strategies across threshold probabilities of approximately 1%-38%. The model also showed greater net benefit than Wagner grade alone below approximately 18%, the range that includes the screening and Youden thresholds. At higher thresholds, the Wagner step curve was greater, reflecting the concentration of test cohort events in Wagner grades 4 and 5; these thresholds correspond to selective prioritization rather than to the screening role proposed for the model.

Table 5. Observed major amputation by prespecified risk band of the primary logistic regression model in the 2022-2025 temporal test cohort (771 patients, 51 events)[a].

Predicted risk band	Patients, n (%)	Observed major amputation, n (%), 95% CI	No major amputation, n (%)
Prespecified risk bands			
<5%	396 (51.4)	0 (0.0), 0.0-1.0	396 (100)
5%-15%	198 (25.7)	14 (7.1), 4.3-11.5	184 (92.9)
15%-30%	123 (16)	15 (12.2), 7.5-19.1	108 (87.8)
≥30%	54 (7)	22 (40.7), 28.7-54.0	32 (59.3)

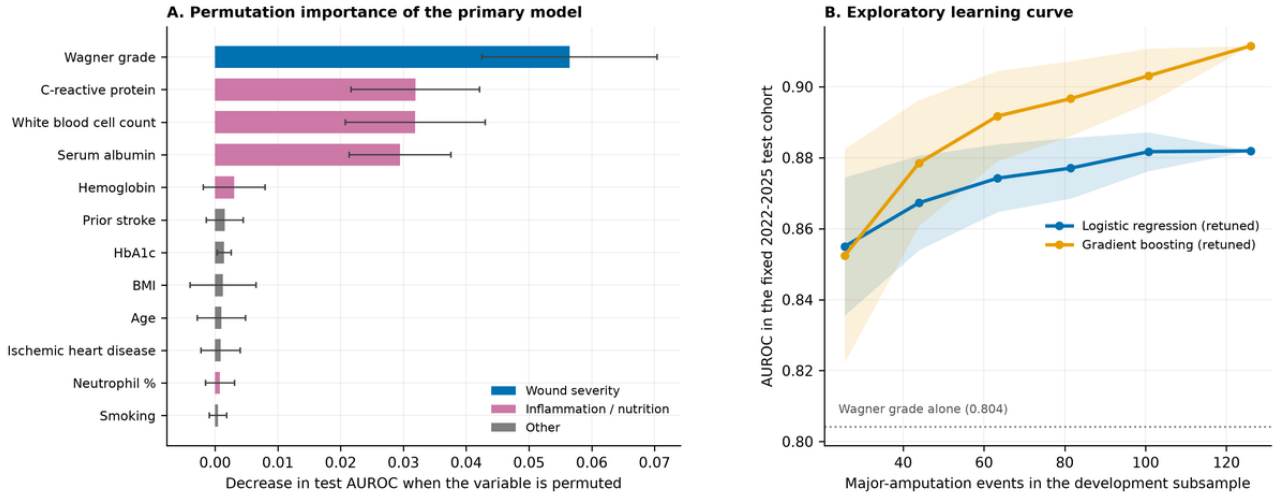
[a]Patients in the "no major amputation" column retained the limb above the ankle; minor amputation may have been performed. CIs are Wilson 95% CIs.

Variable Importance

Permutation importance in the temporal test cohort (Figure 5A) ranked Wagner grade first (mean decrease in AUROC 0.056, SD 0.014), followed by C-reactive protein (0.032, SD 0.010), white blood cell count (0.032, SD 0.011), and serum albumin (0.029, SD 0.008); no other variable exceeded 0.004. The strongest predictors reflected wound severity, systemic inflammation, and physiologic reserve. Standardized coefficients are reported in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 5. Variable importance and exploratory learning curve. (A) Permutation importance of the primary logistic regression model in the temporal test cohort, expressed as the decrease in AUROC when each variable is permuted. Bars show means and error bars show SDs over 50 permutations; the 12 leading variables are displayed, colored by variable domain (wound severity, inflammation or nutrition, and other), and all 20 are reported in Multimedia Appendix 3. (B) AUROC in the fixed 2022-2025 temporal test cohort (51 events) according to the number of major amputation events in the development subsample. Logistic regression and gradient boosting were retuned within each subsample. Points are means over 100 random subsamples at each size below 100%, shaded bands show 1 SD, and the dotted line indicates the Wagner-grade benchmark. The full-cohort end points (126 events) are the single final models reported in Table 2. Because every subsample was evaluated against the same test cohort, the bands describe subsampling variability in this dataset and are not confidence intervals for generalizable performance. AUROC: area under the receiver operating characteristic curve; HbA1c: hemoglobin A1c.



Exploratory Learning Curve

With retuning at each subsample size, gradient boosting showed no advantage over logistic regression in the smallest cohorts, and its mean advantage increased gradually with the number of development events (Figure 5B). At approximately 26 events, mean AUROC was 0.855 for logistic regression and 0.852 for gradient boosting (paired difference –0.003, 95% subsampling interval –0.052 to 0.039; gradient boosting better in 49% of replicates). At approximately 63 events, the difference was +0.017 (–0.008 to 0.036; 92% of replicates), and at approximately 101 events it was +0.021 (0.006-0.033; 99% of replicates). Variability was consistently greater for gradient boosting and greatest at the smallest sample sizes. At the full development cohort of 126 events, the paired test cohort difference was +0.030 (95% CI 0.005-0.059). These results are consistent with the greater data requirements of flexible models [41,42] but do not define a general minimum event count.

Segmental Limb Pressures

All 4 wound-side segmental pressures were available for 1440 of 2171 (66.3%) patients. Availability was related to outcome: the amputation rate was 11.8% (65/550) among the 550 patients (25.3%) with no recorded wound-side pressure vs 6.4% (92/1440) among those with complete measurements, so vascular data missingness cannot be treated as completely at random. The identifiable care-related mechanisms behind the missing measurements are detailed in Multimedia Appendix 2.

To estimate the prognostic contribution of the pressure values without conflating it with their availability, we repeated the comparison in complete cases (944 development patients with 67 events; 496 test patients with 25 events). Adding the 4 pressures did not clearly change discrimination: change in AUROC –0.004 (95% CI –0.044 to 0.035) for gradient boosting and +0.015 (95% CI –0.014 to 0.044) for logistic regression. The CIs remain compatible with modest benefit. These estimates do not demonstrate incremental discrimination in this subset but do not exclude a prognostic role for perfusion. We retained patients with missing pressures in the primary analysis because complete-case restriction would have removed 48.0% (85/177) of all events.

Sensitivity Analyses

After refitting the primary algorithm, the AUROC was 0.891 when in-hospital deaths were excluded, 0.880 when bilateral cases were excluded, and 0.883 when 2020-2021 admissions were excluded; calibration remained similar in the last analysis. For the composite outcome of major amputation or in-hospital death (200 development and 87 test events), the AUROC was 0.870 and the calibration slope was 1.14. Applied unchanged, the primary model score also discriminated in-hospital death alone (46 deaths; AUROC 0.799, 95% CI 0.746-0.847) and major amputation among the 725 patients who survived to discharge (41 events; AUROC 0.890, 95% CI 0.848-0.928).

In the post hoc analyses, refitting without glycated hemoglobin gave an AUROC of 0.883 (95% CI 0.843-0.918; paired difference +0.002, 95% CI –0.004 to 0.008), despite its availability falling from 92.4% to 78.1% between eras. Refitting without albumin gave 0.871 (95% CI 0.828-0.907; paired difference –0.011, 95% CI –0.026 to 0.003), and refitting and evaluating only in patients with an observed albumin value (1005 development patients with 115 events; 593 test patients with 47 events) gave 0.861 (95% CI 0.813-0.902). Adding the presenting wound type to the 20 prespecified variables changed discrimination by +0.005 (95% CI –0.006 to 0.015). These analyses showed no clear incremental discrimination from glycated hemoglobin or presenting wound type in this test cohort; the albumin-restricted analysis changes the evaluated

XSL•FO
RenderX

population and does not exclude bias from informative ordering, and none of these analyses establishes transportability under different measurement practices. When the procedure was repeated in the first eligible admission of every patient (2010 development patients with 147 events; 931 test patients with 58 events), the AUROC was 0.861 (95% CI 0.811-0.905), compared with 0.798 for Wagner grade alone, with a calibration slope of 1.34 and an observed-to-expected ratio of 0.84. Applied unchanged to the first eligible admission of all 931 test-period patients, the primary model had an AUROC of 0.855 (95% CI 0.807-0.899), a calibration slope of 1.15, and an observed-to-expected ratio of 0.65; among the 160 test-period patients with more than 1 hospitalization (7 events), the AUROC was 0.643 (95% CI 0.386-0.884). Full results are reported in Multimedia Appendix 4.

Discussion

Principal Findings

In this internal temporal validation at a single tertiary center, the primary ridge logistic regression model, selected by a prespecified rule and built from 20 routinely collected admission variables, discriminated in-hospital major amputation in a later cohort with an AUROC of 0.882 (95% CI 0.841-0.917), compared with 0.804 for Wagner grade alone, the benchmark universally available at admission. Discrimination, however, was only half of the picture. Absolute risk was overpredicted: mean predicted risk was 9.6% against an observed rate of 6.6%, with a calibration intercept of –0.47, a slope of 1.33, and an observed-to-expected ratio of 0.69. The model's practical contribution was therefore risk stratification rather than reliable absolute probabilities or binary classification: no major amputation occurred among 396 patients in the lowest band, 40.7% of 54 patients in the highest band underwent major amputation, and the screening threshold captured 50 of 51 events at the cost of flagging 44.5% of the cohort. These results support further evaluation of the score for prioritizing the timing of assessment within an existing pathway; they do not support replacing clinical judgment, classifying a limb as salvageable, or transporting the probabilities to another institution without validation.

Appropriate use depends on what the model estimates. Major amputation is a care-dependent outcome shaped by wound extent, infection, perfusion, host reserve, treatment timing, and patient preference; the score estimates the probability of that outcome under 1 institution's pathway (Figure 1) and does not estimate biological limb viability or the causal benefit of amputation, which would require causal methods not used here [29,30]. A high score means that major amputation was frequent among similar patients under the observed pathway, not that amputation was necessary, and 59.3% of patients in the highest band did not undergo major amputation.

Discrimination was maintained in the later era but calibration was not: discrimination was similar after excluding the pandemic-disrupted years (0.883 vs 0.882), whereas the observed-to-expected ratio remained low (0.69 in the primary analysis, 0.67 after that exclusion) and the crude annual event rate declined from 12.8% in 2015 to 5.0% in 2025. These findings suggest a decline in baseline risk that affected calibration more than the ranking of patients; this would not be apparent from the AUROC alone. Consequently, neither the predicted probabilities nor the proposed risk-band boundaries should be transported directly to other institutions: independent external validation should first establish whether discrimination holds and how far calibration departs, and intercept and slope updating would be needed before local use [23-25,32].

Multidisciplinary Clinical Interpretation

For the reconstructive surgeon, a high predicted risk could prompt earlier assessment of source control, tissue viability, and whether a durable soft-tissue envelope can be achieved, determinants of reconstructability that Wagner grade does not capture and that may explain divergent outcomes within a single grade; the high-risk band could prompt early senior review and coordinated staging of debridement, revascularization, and reconstruction, not therapeutic nihilism [8-12]. For the diabetologist, the weight carried by C-reactive protein, white blood cell count, and albumin indicates that systemic inflammation and physiologic reserve add information beyond wound depth. Albumin is a negative acute-phase reactant rather than a nutritional target [43-46], and discrimination was retained when the model was refitted in patients with measured albumin (AUROC 0.861), although that restriction changes the evaluated population and does not exclude bias from informative ordering; the low importance of glycated hemoglobin, which anemia and renal disease confound in this population, does not diminish the role of inpatient glycemic management. For the vascular specialist, the score does not indicate revascularization candidacy and does not replace limb-based hemodynamic and anatomic evaluation [10].

Algorithm Choice and Development Cohort Size

Development cohort cross-validation selected logistic regression; the flexible algorithms appeared stronger only after the test cohort was opened, random forest by 0.033 (95% CI 0.009-0.062), and gradient boosting by 0.030 (95% CI 0.005-0.059). These gains may be genuine, but comparing several algorithms on a single test cohort of 51 events risks selection optimism, and we did not redesignate the primary model; the comparisons are exploratory and descriptive and do not establish the superiority of any algorithm. In the learning-curve analysis, the advantage of boosting grew with event count while its variability was greatest in the smallest samples, but because every subsample was scored against the same 51 test events, the analysis describes this dataset only. Penalized regression therefore remains a stable, interpretable starting point in cohorts of this size, and no universal event-count threshold can be drawn from a single dataset [41,42].

Segmental Limb Pressures

Adding the 4 wound-side segmental pressures did not clearly improve discrimination in complete cases (change in AUROC –0.004, 95% CI –0.044 to 0.035, for gradient boosting; +0.015, 95% CI –0.014 to 0.044, for logistic regression). With 25 test events and selective availability of the measurements, the

XSL•FO
RenderX

estimates are imprecise, and a modest benefit cannot be excluded.

The care pathway likely explains this finding. Vascular assessment was routine, and clinically important ischemia prompted imaging and revascularization; successful revascularization in the patients with the poorest measurements would attenuate the association between admission pressures and amputation, a treatment paradox recognized in clinical prediction [29,30]. A second explanation is medial arterial calcification, which can falsely elevate cuff pressures at any level and is most prevalent in the dialysis-dependent patients in whom measurements were most often missing; the less affected measures, toe pressure and transcutaneous oxygen pressure, were not recorded [33,34]. Neither explanation can be verified without revascularization detail, and neither justifies de-emphasizing vascular evaluation.

Comparison With Prior Work

Our earlier study of 1548 patients with limb-threatening diabetes-related foot ulcers found the Wagner and Texas classifications independently prognostic for amputation or death and, among patients receiving dialysis, found that only C-reactive protein and albumin retained validity across all 4 wound classifications [15]. This study extends those observations: the same constructs (wound grade, C-reactive protein, white blood cell count, and albumin) dominated a model fixed in one era and evaluated in a later one, consistent with meta-analyses and a contemporary review identifying wound severity, infection, inflammatory markers, and renal impairment as major correlates of amputation [47-50].

Published machine learning models for diabetes-related amputation pool to an AUROC near 0.81, with wide variation and little temporal or external validation [18-20]. Our temporally validated AUROC of 0.882, with tree-based models at 0.903 to 0.915, is at the upper end of that range; unlike most prior models, we also report calibration, predictive values, and net benefit, with the model selection rule and thresholds prespecified. Wagner grade served as the benchmark because it is universally available at admission, and it performed identically as a categorical factor or a linear term. The comparison therefore shows that routinely collected admission data add prognostic information beyond Wagner grade alone; it does not show that the model outperforms contemporary multidimensional limb-threat classifications such as WIfI [16,17], against which it was not tested. A full WIfI classification could not be reconstructed: its ischemia axis requires graded pressures or oximetry, which duplex scanning does not provide, and a partial score would have underestimated this stronger comparator.

Clinical Role, Assessment Workload, and Clinical Impact

The intended role of the model is to inform the timing and intensity of assessment within a pathway in which every patient is already assessed, not to decide who receives assessment. At the screening threshold, of 771 patients, 343 (44.5%) would have been flagged to capture 50 of 51 amputations, with a positive predictive value of 14.6%; 293 flagged patients did not undergo major amputation. Escalating review in nearly half of admissions is a substantial demand. Whether a service can absorb it, or whether a more selective threshold such as the top 10% band (75 patients, positive predictive value 34.7%) is a more realistic trigger for senior review and early goals-of-care discussion, will depend on local capacity and on the weight clinicians give to a missed amputation relative to an earlier-than-necessary review. A selective threshold would need separate justification: at the top 10% cutoff the rule identified only half of the amputations (sensitivity 0.510), and at threshold probabilities above approximately 18% the decision curve favored Wagner grade alone (Figure 4B). No threshold should trigger amputation automatically. Because these are retrospective classifications, a flagged patient who kept the limb cannot be equated with an unnecessary assessment, and, conversely, neither the absence of events in the lowest band nor the high negative predictive value justifies delaying vascular evaluation or debridement, since every patient in this cohort received structured multidisciplinary care.

A related and more fundamental question is whether acting on a high predicted risk prevents amputation. Earlier recognition of a high-risk presentation could support timely source control, vascular evaluation and revascularization, and limb-salvage planning, all of which current guidelines recommend [10,12,51]; the evidence for those components of care, however, does not establish the incremental benefit of using this model to prioritize them. Every patient in our cohort was managed within a structured multidisciplinary pathway, and this retrospective study did not compare model-guided care with usual care, so we cannot estimate how many of the 51 amputations, if any, would have been averted by earlier escalation. Net benefit in the decision curve quantifies the trade-off between true- and false-positive classifications at a given threshold [38,40]; it does not demonstrate that the triggered action is effective. That question requires a prospective impact evaluation, ideally a randomized or stepped-wedge comparison of model-guided prioritization against the existing pathway, with assessment timing, resource use, and limb outcomes as end points [31,32].

The next phase should be sequenced. Independent external validation in other centers should come first [23,25], followed by local recalibration where calibration departs, a period of silent prospective evaluation before predictions are shown to clinicians, and then the prospective impact evaluation described in the preceding paragraph. Prospective data collection should record, before definitive treatment, the full WIfI components with a standardized perfusion panel (toe pressure, transcutaneous oxygen pressure, and pedal waveforms), infection severity graded by International Working Group on the Diabetic Foot/Infectious Diseases Society of America criteria [12], prior ulceration and amputation, and complete revascularization and amputation detail.

Limitations

The study has important limitations. The validation was an internal temporal validation: it used later patients from the same tertiary center, treated within the same multidisciplinary and revascularization pathway, in a health system in which podiatry does not exist as a separate profession. It is not an independent

XSL•FO
RenderX

external validation, and differences in case mix, referral patterns, treatment practices, and institutional thresholds for major amputation may all limit transportability. The study was retrospective. Major amputation is care-dependent and partly reflects treatment selection, timing, functional goals, and patient preference, none of which were fully captured; the score also discriminated in-hospital death (AUROC 0.799, 95% CI 0.746-0.847), indicating that it incorporates overall illness severity, and the composite outcome offers a complementary target when death competes with amputation, although it is not a competing-risk estimate of amputation. Eligibility required a single hospitalization over the whole study window, which conditions on the absence of later admissions and excluded 770 of 2941 (26.2%) patients. Because death or major amputation during an index admission makes later foot-related admissions less likely, this restriction enriched the cohort for adverse index outcomes: the 770 excluded patients had a major amputation rate of 3.6% (28/770) and, by construction, no in-hospital deaths at their first admission, compared with 8.2% (177/2171) and 6.6% (144/2171) in the study cohort. Applied unchanged to all first admissions in the test period, the model largely retained discrimination (AUROC 0.855 vs 0.882) but overpredicted risk to a greater extent (observed-to-expected ratio 0.65), and its performance at the first admission of patients who were later readmitted was poor and imprecise (AUROC 0.643, 7 events). Performance in an unselected admission cohort and in recurrent episodes therefore requires separate validation. Because the outcome was ascertained only during the index hospitalization, shorter admissions would lower apparent event rates; median length of stay did not shorten between eras (22 vs 25 days), which makes that explanation unlikely but does not exclude it.

Absolute risk was not adequately calibrated in the later cohort: the observed-to-expected ratio was 0.69, the intercept was –0.47, and the slope was 1.33 (95% CI 1.07-1.67). The test cohort contained 51 amputations, so threshold and risk-band estimates are imprecise, as the CIs in Tables 4 and 5 show; model performance was not examined within patient subgroups such as sex, age, or dialysis status. Laterality and the indication, timing, and outcome of revascularization were unavailable; through-knee and ankle-level (Syme) procedures were not separately coded; segmental limb pressures were selectively missing; no toe pressure or transcutaneous oxygen pressure was recorded, and an ankle-brachial ratio was derivable in only 69.2% (1503/2171) of patients and is itself unreliable in medial arterial calcification; and a full WIfI classification could not be reconstructed. Ulcer duration and dimensions, neuropathy, prior ulcer or amputation, graded infection severity and sepsis physiology, microbiology, diabetes treatment, and functional status were not modeled. No test-cohort event occurred in Wagner grade 0-2, so performance in low-grade disease is uncertain. Calendar-era comparisons were post hoc, the decision curve threshold range was investigator-chosen, and the secondary algorithm comparisons and learning-curve analysis were exploratory. No external or prospective validation has been performed, no impact study has been undertaken, and no clinical calculator is offered.

Conclusions

In an internal temporal validation at a single tertiary center, a ridge logistic regression model, selected by a prespecified rule and based on routinely collected admission data, added discrimination for in-hospital major amputation beyond Wagner grade alone across an 11-year period but overpredicted absolute risk in the later cohort. The score estimates a care-dependent outcome under 1 institution's pathway; this study provides no evidence of superiority over multidimensional classifications such as WIfI, and whether acting on the score improves outcomes is untested. Random forest and gradient boosting showed small secondary gains, and segmental limb pressures showed no clear incremental value in selectively measured complete cases; neither finding alters this message. More broadly, the study illustrates that a model can rank patients well in a later era while its absolute probabilities drift as baseline risk falls, so that discrimination alone is an insufficient basis for deployment. The model's potential role is to support earlier multidisciplinary assessment and prognostic communication, never to determine amputation or to withhold limb-salvage care, and that role must be earned through independent external validation in unselected admission cohorts, recalibration where needed, silent prospective evaluation, and formal assessment of clinical impact and workload.

Acknowledgments

Patients and the public were not involved in the design, conduct, or reporting of this study.

Data Availability

The data supporting the findings of this study contain information that could compromise the privacy of research participants and are therefore not publicly available. A deidentified analytic data dictionary, the model coefficients and hyperparameters (Multimedia Appendix 3), the serialized model objects with their imputer and scaler, and the full analysis code are available from the corresponding author on reasonable request, subject to approval by the Institutional Review Board of the Chang Gung Medical Foundation.

Funding

The authors declared that no financial support was received for this work.

Generative AI Use

During preparation and revision of this work, the authors used generative AI tools (ChatGPT [OpenAI]; Claude [Anthropic]) in 2026 to assist with language and structural editing, with drafting portions of the analysis code, and with organizing the reviewer

XSL•FO
RenderX

comments and drafting revision passages and point-by-point responses. The authors reviewed and revised all AI-assisted text, reviewed the analysis code, and independently specified and verified all analyses, numerical results, references, and conclusions. These tools were not used to generate study data, to perform analyses autonomously, or to make clinical or analytic decisions. The authors take full responsibility for the content of the publication.

Authors' Contributions

Conceptualization: YTF, YYH, RWH
Data curation: YYH
Formal analysis: YTF
Investigation: YYH, JTY
Methodology: YTF, RWH
Project administration: RWH
Resources: JTY
Software: YTF
Supervision: RWH
Visualization: YTF
Writing – original draft: YTF
Writing – review & editing: YTF, YYH, JTY, RWH
All authors read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Completed TRIPOD+AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + Artificial Intelligence) checklist.
[DOCX File , 15 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Missingness by cohort and outcome, and mechanisms of vascular-data missingness.
[DOCX File , 21 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Hyperparameter grids, selected hyperparameters, and primary-model coefficients.
[DOCX File , 20 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Sensitivity analyses, event rates by calendar year, secondary-model results, and patient and event counts by calibration decile.
[DOCX File , 24 KB-Multimedia Appendix 4]

References

1. McDermott K, Fang M, Boulton AJM, Selvin E, Hicks CW. Etiology, epidemiology, and disparities in the burden of diabetic foot ulcers. Diabetes Care. 2023;46(1):209-221. [FREE Full text] [doi: 10.2337/dci22-0043] [Medline: 36548709]
2. Singh N, Armstrong DG, Lipsky BA. Preventing foot ulcers in patients with diabetes. JAMA. 2005;293(2):217-228. [doi: 10.1001/jama.293.2.217] [Medline: 15644549]
3. Hicks CW, Canner JK, Mathioudakis N, Lippincott C, Sherman RL, Abularrage CJ. Incidence and risk factors associated with ulcer recurrence among patients with diabetic foot ulcers treated in a multidisciplinary setting. J Surg Res. 2020;246:243-250. [doi: 10.1016/j.jss.2019.09.025] [Medline: 31610352]
4. Armstrong DG, Tan TW, Boulton AJM, Bus SA. Diabetic foot ulcers: a review. JAMA. 2023;330(1):62-75. [doi: 10.1001/jama.2023.10578] [Medline: 37395769]
5. Armstrong DG, Boulton AJM, Bus SA. Diabetic foot ulcers and their recurrence. N Engl J Med. 2017;376(24):2367-2375. [doi: 10.1056/NEJMra1615439] [Medline: 28614678]
6. Armstrong DG, Swerdlow MA, Armstrong AA, Conte MS, Padula WV, Bus SA. Five year mortality and direct costs of care for people with diabetic foot complications are comparable to cancer. J Foot Ankle Res. 2020;13(1):16. [FREE Full text] [doi: 10.1186/s13047-020-00383-2] [Medline: 32209136]

XSL•FO
RenderX

7. Wukich DK, Raspovic KM, Suder NC. Patients with diabetic foot disease fear major lower-extremity amputation more than death. Foot Ankle Spec. 2018;11(1):17-21. [doi: 10.1177/1938640017694722] [Medline: 28817962]
8. Singer AJ, Tassiopoulos A, Kirsner RS. Evaluation and management of lower-extremity ulcers. N Engl J Med. 2017;377(16):1559-1567. [doi: 10.1056/NEJMra1615243] [Medline: 29045216]
9. Everett E, Mathioudakis N. Update on management of diabetic foot ulcers. Ann N Y Acad Sci. 2018;1411(1):153-165. [FREE Full text] [doi: 10.1111/nyas.13569] [Medline: 29377202]
10. Fitridge R, Chuter V, Mills J, Hinchliffe R, Azuma N, Behrendt CA, et al. The intersocietal IWGDF, ESVS, SVS guidelines on peripheral artery disease in people with diabetes mellitus and a foot ulcer. J Vasc Surg. 2023;78(5):1101-1131. [FREE Full text] [doi: 10.1016/j.jvs.2023.07.020] [Medline: 37724985]
11. Aragón-Sánchez J, Lázaro-Martínez JL, Pulido-Duque J, Maynar M. From the diabetic foot ulcer and beyond: how do foot infections spread in patients with diabetes? Diabet Foot Ankle. 2012;3:18693. [FREE Full text] [doi: 10.3402/dfa.v3i0.18693] [Medline: 23050067]
12. Senneville É, Albalawi Z, van Asten SA, Abbas ZG, Allison G, Aragón-Sánchez J, et al. IWGDF/IDSA guidelines on the diagnosis and treatment of diabetes-related foot infections (IWGDF/IDSA 2023). Diabetes Metab Res Rev. 2024;40(3):e3687. [doi: 10.1002/dmrr.3687] [Medline: 37779323]
13. Frykberg RG, Zgonis T, Armstrong DG, Driver VR, Giurini JM, Kravitz SR, et al. American College of Foot and Ankle Surgeons. Diabetic foot disorders: a clinical practice guideline (2006 revision). J Foot Ankle Surg. 2006;45(5 Suppl):S1-S66. [doi: 10.1016/S1067-2516(07)60001-5] [Medline: 17280936]
14. Alahakoon C, Fernando M, Galappaththy C, Lazzarini P, Moxon JV, Jones R, et al. Repeatability, completion time, and predictive ability of four diabetes-related foot ulcer classification systems. J Diabetes Sci Technol. 2023;17(1):35-41. [FREE Full text] [doi: 10.1177/1932296820986548] [Medline: 33451251]
15. Chang YC, Huang YY, Hung SY, Yeh JT, Lin CW, Chen IW, et al. Are current wound classifications valid for predicting prognosis in people treated for limb-threatening diabetic foot ulcers? Int Wound J. 2024;21(1):e14338. [FREE Full text] [doi: 10.1111/iwj.14338] [Medline: 37555265]
16. Mills JL, Conte MS, Armstrong DG, Pomposelli FB, Schanzer A, Sidawy AN, et al. Society for Vascular Surgery Lower Extremity Guidelines Committee. The Society for Vascular Surgery Lower Extremity Threatened Limb Classification System: risk stratification based on wound, ischemia, and foot infection (WIfI). J Vasc Surg. 2014;59(1):220-234.e2. [FREE Full text] [doi: 10.1016/j.jvs.2013.08.003] [Medline: 24126108]
17. Monteiro-Soares M, Hamilton EJ, Russell DA, Srisawasdi G, Boyko EJ, Mills JL, et al. Guidelines on the classification of foot ulcers in people with diabetes (IWGDF 2023 update). Diabetes Metab Res Rev. 2024;40(3):e3648. [doi: 10.1002/dmrr.3648] [Medline: 37179483]
18. Chen Z, Liu X, Li S, Wu Z, Tan H, Yu F, et al. Machine learning for the prediction of diabetes-related amputation: a systematic review and meta-analysis of diagnostic test accuracy. Clin Exp Med. 2025;25(1):151. [doi: 10.1007/s10238-025-01697-w] [Medline: 40348887]
19. Stefanopoulos S, Qiu Q, Ren G, Ahmed A, Osman M, Brunicardi FC, et al. A machine learning model for prediction of amputation in diabetics. J Diabetes Sci Technol. 2024;18(4):874-881. [doi: 10.1177/19322968221142899] [Medline: 36476059]
20. Xie P, Li Y, Deng B, Du C, Rui S, Deng W, et al. An explainable machine learning model for predicting in-hospital amputation rate of patients with diabetic foot ulcer. Int Wound J. 2022;19(4):910-918. [FREE Full text] [doi: 10.1111/iwj.13691] [Medline: 34520110]
21. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605. [doi: 10.1136/bmj.b605] [Medline: 19477892]
22. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2021;14(1):49-58. [FREE Full text] [doi: 10.1093/ckj/sfaa188] [Medline: 33564405]
23. Riley RD, Archer L, Snell KIE, Ensor J, Dhiman P, Martin GP, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. BMJ. 2024;384:e074820. [FREE Full text] [doi: 10.1136/bmj-2023-074820] [Medline: 38224968]
24. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. [FREE Full text] [doi: 10.1186/s12916-019-1466-7] [Medline: 31842878]
25. Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. BMC Med. 2023;21(1):70. [FREE Full text] [doi: 10.1186/s12916-023-02779-w] [Medline: 36829188]
26. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. [FREE Full text] [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
27. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388:e082505. [doi: 10.1136/bmj-2024-082505] [Medline: 40127903]

XSL•FO
RenderX

28. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ. 2024;384:e074819. [FREE Full text] [doi: 10.1136/bmj-2023-074819] [Medline: 38191193]
29. Sperrin M, Diaz-Ordaz K, Pajouheshnia R. Invited commentary: treatment drop-in—making the case for causal prediction. Am J Epidemiol. 2021;190(10):2015-2018. [doi: 10.1093/aje/kwab030] [Medline: 33595073]
30. Lin L, Sperrin M, Jenkins DA, Martin GP, Peek N. A scoping review of causal methods enabling predictions under hypothetical interventions. Diagn Progn Res. 2021;5(1):3. [FREE Full text] [doi: 10.1186/s41512-021-00092-9] [Medline: 33536082]
31. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagn Progn Res. 2018;2:11. [FREE Full text] [doi: 10.1186/s41512-018-0033-6] [Medline: 31093561]
32. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-698. [doi: 10.1136/heartjnl-2011-301247] [Medline: 22397946]
33. Chuter V, Schaper N, Mills J, Hinchliffe R, Russell D, Azuma N, et al. Effectiveness of bedside investigations to diagnose peripheral artery disease among people with diabetes mellitus: a systematic review. Diabetes Metab Res Rev. 2024;40(3):e3683. [FREE Full text] [doi: 10.1002/dmrr.3683] [Medline: 37477087]
34. Lanzer P, Hannan FM, Lanzer JD, Janzen J, Raggi P, Furniss D, et al. Medial arterial calcification: JACC state-of-the-art review. J Am Coll Cardiol. 2021;78(11):1145-1165. [FREE Full text] [doi: 10.1016/j.jacc.2021.06.049] [Medline: 34503684]
35. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441. [doi: 10.1136/bmj.m441] [Medline: 32188600]
36. Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE, Moons KGM, et al. Minimum sample size for developing a multivariable prediction model: PART II—binary and time-to-event outcomes. Stat Med. 2019;38(7):1276-1296. [FREE Full text] [doi: 10.1002/sim.7992] [Medline: 30357870]
37. van den Goorbergh R, van Smeden M, Timmerman D, Van Calster B. The harm of class imbalance corrections for risk prediction models: illustration and simulation using logistic regression. J Am Med Inform Assoc. 2022;29(9):1525-1534. [FREE Full text] [doi: 10.1093/jamia/ocac093] [Medline: 35686364]
38. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-574. [FREE Full text] [doi: 10.1177/0272989X06295361] [Medline: 17099194]
39. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352:i6. [FREE Full text] [doi: 10.1136/bmj.i6] [Medline: 26810254]
40. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019;3:18. [FREE Full text] [doi: 10.1186/s41512-019-0064-7] [Medline: 31592444]
41. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12-22. [doi: 10.1016/j.jclinepi.2019.02.004] [Medline: 30763612]
42. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol. 2014;14:137. [FREE Full text] [doi: 10.1186/1471-2288-14-137] [Medline: 25532820]
43. Evans DC, Corkins MR, Malone A, Miller S, Mogensen KM, Guenter P, et al. ASPEN Malnutrition Committee. The use of visceral proteins as nutrition markers: an ASPEN position paper. Nutr Clin Pract. 2021;36(1):22-28. [doi: 10.1002/ncp.10588] [Medline: 33125793]
44. Ishida S, Hashimoto I, Seike T, Abe Y, Nakaya Y, Nakanishi H. Serum albumin levels correlate with inflammation rather than nutrition supply in burns patients: a retrospective study. J Med Invest. 2014;61(3-4):361-368. [FREE Full text] [doi: 10.2152/jmi.61.361] [Medline: 25264055]
45. Eckart A, Struja T, Kutz A, Baumgartner A, Baumgartner T, Zurfluh S, et al. Relationship of nutritional status, inflammation, and serum albumin levels during acute illness: a prospective study. Am J Med. 2020;133(6):713-722.e7. [FREE Full text] [doi: 10.1016/j.amjmed.2019.10.031] [Medline: 31751531]
46. Li Y, Chen L, Yang X, Cui H, Li Z, Chen W, et al. Dynamic association of serum albumin changes with inflammation, nutritional status and clinical outcomes: a secondary analysis of a large prospective observational cohort study. Eur J Med Res. 2025;30(1):679. [FREE Full text] [doi: 10.1186/s40001-025-02925-5] [Medline: 40717075]
47. Lin C, Liu J, Sun H. Risk factors for lower extremity amputation in patients with diabetic foot ulcers: a meta-analysis. PLoS One. 2020;15(9):e0239236. [FREE Full text] [doi: 10.1371/journal.pone.0239236] [Medline: 32936828]
48. Sen P, Demirdal T, Emir B. Meta-analysis of risk factors for amputation in diabetic foot infections. Diabetes Metab Res Rev. 2019;35(7):e3165. [doi: 10.1002/dmrr.3165] [Medline: 30953392]
49. Luo Y, Liu C, Li C, Jin M, Pi L, Jin Z. The incidence of lower extremity amputation and its associated risk factors in patients with diabetic foot ulcers: a meta-analysis. Int Wound J. 2024;21(7):e14931. [doi: 10.1111/iwj.14931] [Medline: 38972836]
50. Jeffcoate W, Boyko EJ, Game F, Cowled P, Senneville E, Fitridge R. Causes, prevention, and management of diabetes-related foot ulcers. Lancet Diabetes Endocrinol. 2024;12(7):472-482. [doi: 10.1016/S2213-8587(24)00110-4] [Medline: 38824929]

XSL•FO
RenderX

51. Schaper NC, van Netten JJ, Apelqvist J, Bus SA, Fitridge R, Game F, et al. IWGDF Editorial Board. Practical guidelines on the prevention and management of diabetes-related foot disease (IWGDF 2023 update). Diabetes Metab Res Rev. 2024;40(3):e3657. [FREE Full text] [doi: 10.1002/dmrr.3657] [Medline: 37243927]

Abbreviations

AUPRC: area under the precision-recall curve
AUROC: area under the receiver operating characteristic curve
SMD: standardized mean difference
TRIPOD+AI: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + AI
WIfI: Wound, Ischemia, and foot Infection

Edited by I Steenstra; submitted 29.Aug.2026; peer-reviewed by M Yovera-Aldana, S Kaluva; comments to author 01.Oct.2026; revised version received 07.Oct.2026; accepted 08.Oct.2026; published 13.Nov.2026

Please cite as:
Fu Y-T, Huang Y-Y, Yeh J-T, Huang R-W
Admission-Time Prediction of In-Hospital Major Amputation in Diabetes-Related Foot Disease: Model Development and Temporal Validation Study
JMIR Med Inform 2026;14:e110740
URL: https://medinform.jmir.org/2026/1/e110740
doi: 10.2196/110740
PMID:

©Yu-Tung Fu, Yu-Yao Huang, Jiun-Ting Yeh, Ren-Wen Huang. Originally published in JMIR Medical Informatics (https://medinform.jmir.org), 13.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on https://medinform.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX