Zaina Mirza[1], Nila N. Sari[1, 2], Rajna Ogrin[3, 4], Samantha Hanna[5], Kate Waller[6], Barbara Polus[1], Elif Ekinci[4, 7, 8], and Dinesh K. Kumar[1, 4]

Feasibility of Utilizing Routine Clinical Data for DFU Healing Prognosis in Resource-Limited Settings: A Pilot Study.

ORCIDs
Zaina Mirza: 0009-0001-0449-7574
Nila Novita Sari: 0000-0002-9892-9912
Rajna Ogrin: 0000-0002-4192-7254
Samantha Hanna: 0009-0001-5598-0224
Kate Waller: 0000-0002-5911-9626
Barbara Polus: 0000-0003-2227-2659
Elif Ekinci: 0000-0003-2372-395X
Dinesh K. Kumar: 0000-0003-3602-4023

[1] *School of Engineering, RMIT University, Victoria, Australia*
[2] *Electrical Engineering, Politeknik Negeri Bandung, Indonesia*
[3] *Bolton Clarke Research Institute, Melbourne, VIC, Australia*
[4] *Australian Centre for Accelerating Diabetes Innovations, Melbourne Medical School, University of Melbourne, Victoria, Australia*
[5] *Department of Podiatry, Austin Hospital, Victoria, Australia*
[6] *Department of Podiatry and High Risk Foot Service, St Vincent's Hospital, Melbourne, Victoria, Australia*
[7] *Department of Endocrinology, Austin Health, Melbourne, Victoria, Australia*
[8] *Department of Medicine, Melbourne Medical School, University of Melbourne, Victoria, Australia*

Abstract – Diabetic foot ulcer (DFU) healing prognosis is clinically important for identifying high-risk patients who require earlier care escalation. This study analyzed healing outcome prediction in 51 patients with a DFU using six routinely collected clinical variables previously associated with non-healing. Most DFU research has focused on ulcer detection rather than healing outcome prediction. Furthermore, existing literature on prognosis often lacks calibration and interpretability, frequently relying on multimodal imaging-based approaches rather than demographic variables alone. Logistic regression established a prognostic baseline (ROC-AUC 0.66), suggesting that routine demographic variables alone provide limited discriminative power without wound-specific characteristics as input. Incorporating HbA1c using a missingness indicator improved sensitivity and classification balance, indicating that clinical data availability itself may carry prognostic information. These findings quantify the performance constraints of clinical datasets, defining the significant gap that future multimodal approaches must fill.

Clinical relevance – The findings demonstrate that routine clinical variables, while limited in isolation, provide an immediate signal for risk stratification. This establishes a preliminary prognostic baseline, offering a feasible risk-assessment tool for regional settings where advanced imaging resources are unavailable.

I. Introduction

Diabetes mellitus is among the most common metabolic diseases. In 2019, 9.3% (463 million) of individuals were affected worldwide [1]. Annually, 18.6 million people who have diabetes are diagnosed with DFUs, leading to debilitating complications including morbidity, lower limb amputations, and mortality [2]. Individuals with DFUs have a five-year mortality rate of 30%, which increases to 70% following amputation [2].

Despite growing research on models for diabetic foot ulcer management, several focus on ulcer detection or risk prediction rather than healing prognosis. Moreover, existing DFU prognostic models rely on wound-specific characteristics or longitudinal wound assessments rather than demographic or routinely collected clinical variables alone, limiting their application at first presentation prior to wound evaluation (*Fig. 1*) [3]. Furthermore, several prognostic models rely on large electronic health records, often necessitating extensive computing infrastructure and resources for development and maintenance [4].

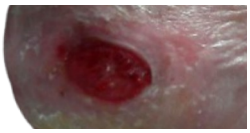

Fig. 1 Representative image of a DFU illustrating ulcer characteristics assessed during routine clinical examination

Berezo et al. utilized extensive electronic health records to classify healing vs non-healing ulcers achieving AU-ROC values of 0.85 to 0.87 [4]. While regression-based analyses on small, clinic-level datasets report modest but interpretable results, few have addressed calibration, which is essential to ensure that model outputs are not artificially extreme.

Imaging-based approaches include RGB clinical photography and planimetric wound area measurement, as well as more advanced modalities such as thermal imaging to quantify wound characteristics. Aliahmad et al. demonstrated the potential of thermal imaging for predicting DFU healing during a four-week follow-up period; however, this approach and imaging-based methods more broadly are constrained by specialized equipment and sensitive to data variability, limiting their scalability in routine practice [5, 6]. These constraints disproportionately impact remote and regional settings, including First Nations communities where access to diabetic foot management services may be limited [7]. In contrast, prognostic models using routine biomarkers can be applied at the first clinical presentation without additional hardware.

This research investigates the feasibility of calibrated prognostic models for DFU healing by proposing a proof-of-concept framework based on standard clinical data. Unlike opaque, black-box machine learning models, the output will include a reproducible and transparent baseline and will demonstrate the prognostic value of informative missingness in routine clinical data. By enabling early identification of non-healing DFUs at baseline assessment, this approach aligns with preventative strategies in diabetic foot management.

II. METHODS

A. Dataset and Outcome Definition

This study was conducted in accordance with approvals from the Austin Health Human Research Ethics Committee (HREC/92993/Austin-2023) and St Vincent's Hospital Melbourne (2024/PID00243). Patient data were collected on REDCap. The study utilized a convenience cohort of n=51 participants, to establish a baseline for foundational data that is available in regional settings.

Raw dataset predictors included gender (male/female), diabetes duration (diagnosis year or elapsed year), types of diabetes treatment (oral hypoglycemic agents, insulin therapy, diet-controlled, other/unknown), HbA1c (% or mmol/mol), age, and smoking status (current, past, never, not stated). The target variable was healing status, which was encoded as 1 (healed) and 0 (non-healed).

B. Data Preparation and Missingness Handling

Firstly, diabetes duration units were standardized to diabetes duration in years, and HbA1c values were converted to mmol/mol. Following this, healing status included amputation, which was grouped with the non-healed outcome (0).

The dataset exhibited class imbalance, particularly in healing status (34 healed vs 17 non-healed), gender (40 male vs 11 female) and incomplete data in selected clinical variables. HbA1c had the highest proportion of missing data (49%) and treatment type was missing in 3.9% of cases. To evaluate the optimal imputation strategy for these variables, we assessed the missingness mechanism as missing completely at random (MCAR) or missing at random (MAR) by testing associations between missingness and observed covariates for HbA1c. Diabetes treatment type was examined descriptively due to the small proportion of missing data.

Based on the analysis, small treatment categories (N/A, diet-controlled, and other; $n \leq 4$) were collapsed into "other/unknown" groups to improve model stability. Moreover, HbA1c missingness patterns were suggestive of MAR, where missingness is associated with observed features. Accordingly, the imputation strategy used for HbA1c was median imputation with an explicit missingness indicator. This allowed patient data to be retained while reducing complete-case bias.

C. Feature Selection

Feature selection was used to empirically prioritize variables with the most stable signal and to reduce model instability. Final performance comparison models were conducted within cross-validated pipelines.

TABLE I. UNIVARIATE ASSOCIATIONS WITH HEALING OUTCOME

Variable	Test	p	Effect size
HbA1c	Welch t-test	0.598	d = 0.24
Age	Welch t-test	0.951	d = −0.02
Diabetes duration	Welch t-test	0.243	d = 0.31
Gender	χ^2 / Fisher	0.185 / 0.147	V = 0.19
Smoking status	χ^2	0.488	V = 0.22
Treatment	χ^2	0.140	V = 0.28

Table I details the univariate associations. Continuous variables (age, HbA1c, diabetes duration) approximated normal distributions and were therefore compared using Welch's t-tests. Results showed small effect sizes with no statistically significant differences between healed and non-healed groups. Categorical variables (gender, smoking, treatment) exhibited larger effects, notably treatment type, followed by smoking status and gender. Therefore, for evaluation categorical variables were grouped for testing, followed by continuous variables, and the full feature set.

D. Model Development and Evaluation

Given the low events-per variable ratio (EPV < 10), modelling was restricted to logistic regression with penalized (L2) and classifiers were trained and tuned using GridSearchCV. This decision prioritizes the stability of the model and interpretability over complex non-linear interactions, which may overfit and demonstrate high variance in small datasets. Categorical variables were one-hot encoded and numeric variables were standardized within a preprocessing pipeline to prevent leakage. Performance was evaluated using 1000 iterations of bootstrap validation to generate 95% confidence intervals.

Due to the small sample size and class imbalance, discrimination and calibration metrics were emphasized. ROC-AUC was used as the primary discrimination metric. Sensitivity and specificity were computed by converting predicted probabilities to class labels using a fixed threshold of 0.5 for comparability across models. Probabilistic calibration was assessed using the Brier score, which characterizes agreement between predicted probabilities and observed outcomes.

1) Incorporation of HbA1c

HbA1c was incorporated into the prognostic pipeline using median imputation and a binary missingness flag, preserving information encoded in the missingness pattern. For modeling, both the HbA1c imputed value and missingness indicator were included as predictors alongside the full set of routinely collected clinical data (n=51). Modeling was restricted to logistic regression and employed the same preprocessing pipeline described above.

III. Results

A. Model Performance Across Feature Sets

TABLE II. Logistic regression performance across feature sets

Feature Set	ROC-AUC (95% CI)	Sensitivity	Specificity	Brier Score (95% CI)
Categorical	**0.66 (0.48–0.82)**	**0.71**	**0.47**	**0.22 (0.20–0.25)**
Cat. + Duration	0.62 (0.44–0.78)	0.62	0.53	0.23 (0.20–0.26)
All Clinical Features (n=27)	0.41 (0.12–0.72)	0.39	0.33	0.28 (0.24–0.33)

TABLE III. Logistic Regression odds ratios for categorical variables

Predictor	Odds Ratio (exp β)
Treatment: Oral (vs insulin)	0.441
Gender: Male (vs female)	2.083
Smoking: Non-smoker (vs current)	1.767

Table II summarizes model performance across all feature sets. Logistic regression demonstrated the strongest performance in the categorical feature space; it achieved comparatively superior discrimination (ROC-AUC = 0.66) and exhibited reasonable calibration (Brier Score = 0.22). This performance was driven by key discriminatory features; patients on oral hypoglycemics were less likely to heal compared to those in the insulin reference group, while males and non-smoking status were associated with higher odds of healing (Table III).

Consistent with univariate findings, adding diabetes duration provided limited predictive value. Moreover, the full feature set relied on complete case data, which reduced the sample size to n=27. This reduction of statistical power degraded performance (ROC-AUC = 0.41), highlighting the limitations of excluding patients due to incomplete medical records.

B. Logistic Regression Discrimination and Calibration

Fig. 1 and *Fig. 2* illustrate the discrimination and calibration performance of the logistic regression model using categorical features.

Fig. 2 Receiver Operating Characteristic (ROC) Curve for logistic regression using categorical features

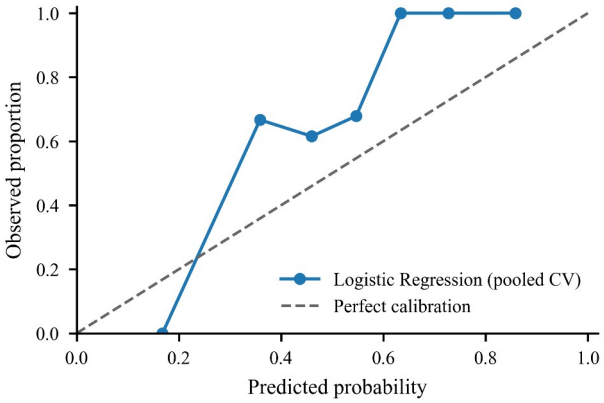

Fig. 3 Calibration curve for logistic regression using categorical features

The ROC curve (*Fig. 2)* demonstrated that the logistic regression model has the capacity to discriminate above random chance using categorical features, supporting its ability to differentiate between healed and non-healed classes. The calibration curve (*Fig. 3*) reflects reasonable probabilistic agreement, although some instability is expected given the small sample size.

C. HbA1c Integration in the Complete Case Dataset

TABLE IV. Effects of HbA1c integration in the full clinical model

Model	ROC-AUC	Sensitivity	Specificity	Brier
Full model + HbA1c (median + missing flag)	0.674	0.824	0.588	0.226

TABLE V. HbA1c-related odds ratios in the full clinical model

Predictor	Odds Ratio
HbA1c missing indicator	**0.396**

Table IV demonstrates model performance when HbA1c was incorporated using median imputation and an explicit missingness indicator. At the fixed 0.5 threshold, both sensitivity and specificity increased, while maintaining calibration suggesting more balanced clinical risk stratification.

Table V demonstrates that the missingness of HbA1c may carry prognostic value, suggesting that the testing behavior itself may encode information related to patient engagement and disease severity.

IV. Discussion

Using a limited feature set (smoking status, gender, and treatment type), the logistic regression model provided the most stable and well-calibrated performance for predicting DFU healing. This stability is likely attributable to the small,

imbalanced dataset. Unlike more flexible, non-linear models, logistic regression learns from a single, global decision boundary, allowing for more stable predictions. Under class imbalance, complex models may yield misleading classification scores, whereas L2 regularization in logistic regression promotes better-calibrated probability estimates. Furthermore, as the aim of this work is to provide reliable probabilistic outputs for DFU healing classification, overconfident predictions pose clinical concerns, as underestimating uncertainty may increase the risk of adverse consequences.

Interpretability allows clinicians to scrutinize the model's output, acting as a transparent decision-support tool rather than a black box. The observed associations with gender may reflect the data imbalance rather than biological differences. Moreover, the apparent influence of non-smokers versus smokers likely reflects behavioral or care-related factors rather than primary causal drivers. Therefore, these associations should not be interpreted as causal because they may reflect confounding by indication and care context such as patient severity and care pathways. Interpretability supports clinical plausibility; however, findings must be assessed cautiously due to the limited and unbalanced sample size.

The findings suggest that the absence of HbA1c testing carries prognostic information. Rather than reflecting glycemic control directly, the association of missing HbA1c and healing outcomes suggests that incomplete medical data may act as a surrogate marker for systemic discontinuity of care or reduced patient engagement, reflecting known risk factors for poor DFU healing. This strengthens the clinical plausibility that data completeness itself may encode relevant clinical signals, particularly in care settings.

Applications include early risk stratification, as this model does not rely on wound characteristics or longitudinal data that may not be available at first presentation. In addition, this framework can be applied in resource-limited settings, such as in regional clinics, where access to specialist referrals may be constrained.

Limitations include the small sample size, class imbalance, and missingness. As this work focused on routinely collected clinical data, other features such as RGB clinical images could better characterize ulcer severity and enhance prognostic modeling. To support improvements to the model, future work should include validation of a larger sample size with a broader feature set.

V. CONCLUSION

In conclusion, this study demonstrates the feasibility of a proof-of-concept approach for predicting diabetic foot ulcer healing outcomes using routinely collected clinical data. Using a limited feature set (gender, smoking status, treatment), logistic regression identified a preliminary signal for discrimination (ROC-AUC = 0.66) while maintaining calibration and interpretability. As these variables are available at first presentation, this approach may support early risk stratification in resource-limited and regional settings such as First Nations communities, where access to specialist services may be limited. The findings further identified data incompleteness as a surrogate marker for discontinuity of care. Collectively, the results demonstrate that small, low-cost clinical datasets can yield baseline discriminative performance when supported by appropriate calibration, providing a foundation for an accessible and stable prognostic support tool for preventative diabetic foot management.

VI. ACKNOWLEDGEMENTS

The authors acknowledge the Australasian Centre for Aboriginal Diabetes and the Australian Podiatry Association.

VII. REFERENCES

[1] P. Saeedi *et al.*, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition," *Diabetes Research and Clinical Practice,* vol. 157, p. 107843, 2019, doi: 10.1016/j.diabres.2019.107843.

[2] D. G. Armstrong, T.-W. Tan, A. J. M. Boulton, and S. A. Bus, "Diabetic Foot Ulcers," *JAMA,* vol. 330, no. 1, pp. 62–75, 2023, doi: 10.1001/jama.2023.10578.

[3] K. Jung *et al.*, "Rapid identification of slow healing wounds," *Wound Repair Regen,* vol. 24, no. 1, pp. 181–188, 2016, doi: 10.1111/wrr.12384.

[4] M. Berezo, J. Budman, D. Deutscher, C. T. Hess, K. Smith, and D. Hayes, "Predicting Chronic Wound Healing Time Using Machine Learning," *Adv. Wound Care,* vol. 11, no. 6, pp. 281–296, 2022, doi: 10.1089/wound.2021.0073.

[5] B. Aliahmad *et al.*, "Is Thermal Imaging a Useful Predictor of the Healing Status of Diabetes-Related Foot Ulcers? A Pilot Study," *J. Diabetes Sci. Technol.,* vol. 13, no. 3, pp. 561–567, May 2019, doi: 10.1177/1932296818803115.

[6] N. N. Sari *et al.*, "Non-invasive imaging techniques for predicting healing status of diabetic foot ulcers: A ten-year systematic review," *Frontiers in Medical Technology,* vol. 7, p. 1648973, 2025, doi: 10.3389/fmedt.2025.

[7] Y. K. Soonarane, M. Kirk, G. Khandaker, and R. Varrall, "Epidemiology and healthcare access inequities in diabetic foot disease: a retrospective study in Central Queensland, Australia," *BMJ Open,* vol. 15, no. 7, p. e098999, 2025, doi: 10.1136/bmjopen-2025-098999.