The International Journal of Medical Science and Health Research

Prognostic Performance of Artificial Intelligence Models in Predicting 12-Week Healing of Chronic Wounds: A Systematic Review And Meta-Analysis

[1] Jeany Thalia Hartono, [2] Fanny Evasari Lesmanawati, [3] Rosalyn Devina Santoso

[1] School of Medicine, Universitas Ciputra Surabaya, Surabaya, Indonesia

[2,3] Department of Plastic, Reconstructive and Aesthetic Surgery, School of Medicine, Universitas Ciputra Surabaya, Surabaya, Indonesia

Corresponding Email : jeanythalia@gmail.com

Article History :

Received date	: 2025/09/09
Revised date	: 2025/10/14
Accepted date	: 2025/11/21
Published date	: 2025/12/18

E-ISSN :
ISSN 3048-1368

P-ISSN
ISSN 3048-1376

ABSTRACT

Introduction

Chronic wounds pose a substantial health burden, requiring intensive, long-term management and carrying a high risk of debilitating complications. Accurate prognosis regarding the probability and rate of healing within the critical 12-week timeframe is essential for optimizing specific care strategies and ensuring the effective allocation of scarce medical resources. Artificial Intelligence (AI), particularly through its foundation in Machine Learning (ML), offers significant potential to enhance prognostic accuracy by rigorously processing vast quantities of Electronic Medical Record (EMR) data and advanced wound imagery.

Methods

This systematic review and meta-analysis was conducted in strict

adherence to the PRISMA 2020 reporting guidelines. Included studies specifically evaluated AI models designed to predict chronic wound healing outcomes within 12 weeks. The methodological quality of these studies was critically assessed using the specialized **Prediction Model Risk of Bias Assessment Tool for Artificial Intelligence (PROBAST+AI)**. Quantitative synthesis was executed to determine the pooled discrimination performance metric, the Area Under the Curve (AUC), and to measure the independent effects of key predictors using pooled Hazard Ratios (HR).

Results

The analysis incorporated two large-scale primary studies boasting high data volumes, alongside several supporting methodological studies. The resultant pooled AUC for AI models reached **0.805 (95% CI: 0.778–0.832)**, definitively confirming significant prognostic capability. Specifically, models utilizing Gradient-Boosted Decision Tree (GBDT) algorithms achieved an AUC of 0.853, a performance level that substantially outperformed conventional Logistic Regression models (AUC 0.712). Assessment utilizing PROBAST+AI consistently highlighted systemic methodological quality issues, predominantly stemming from weak internal validation within the Analysis Domain, which consequently elevated the Overall Risk of Bias. The pooled HR analysis, synthesizing data for 10 critical prognostic predictors, confirmed that local wound characteristics are the paramount determinants of prognosis. **High-Grade Wound Depth (Stage 3/4)** was identified as the single strongest inhibitor of healing (HR **0.65 (95% CI: 0.59–0.71)**), whereas **Normal/Good Vascularization Status** represented the strongest accelerator (HR

1.30 (95% CI: 1.22–1.39)).

Discussion and Conclusion

The prognostic performance demonstrated by AI models is statistically significant and definitively exceeds that of conventional statistical methods. This heightened accuracy is attributed to the inherent capacity of non-linear models to effectively capture complex multi-variable interactions central to wound healing pathophysiology. Notwithstanding the encouraging performance metrics, the documented high risk of *overfitting* due to analytical bias necessitates strict and rigorous external validation prior to any extensive clinical implementation.

Keywords: Artificial Intelligence, Chronic Wounds, Prognosis, 12-Week Healing, Hazard Ratio, AUC, PROBAST+AI, Machine Learning.

INTRODUCTION

Clinical and Technological Background

Clinical Imperative

Chronic wounds, encompassing conditions such as diabetic ulcers and pressure ulcers, impose a protracted and intensive management necessity, leading to profound negative implications for patient quality of life and massive expenditures within healthcare systems (Anon, 2024). The achievement of an accurate prognosis regarding the probability and speed of healing is a critically urgent clinical need, particularly when focused on the initial 12-week period, which is universally recognized as the decisive window for assessing therapeutic success and modifying treatment plans (Cho et al., 2020).

Deficiencies of Traditional Prognostic Methods

Historically, prognostic assessments have been fundamentally dependent upon subjective clinical experience and the utilization of rudimentary scoring systems. The intrinsic limitations of these conventional methods have consistently led to prognostic outcomes characterized by suboptimal accuracy, complicating effective clinical decision-making.

The Role of Artificial Intelligence

The technological proliferation of Artificial Intelligence (AI) provides an innovative and promising solution designed to significantly elevate both the diagnostic and prognostic precision inherent in modern wound management (Anon, 2023). AI, specifically utilizing sophisticated algorithms embedded in Machine Learning (ML) and Deep Learning (DL), possesses the unique capability to process, filter, and extract meaningful patterns from vast volumes of Electronic Medical Record (EMR) data. This comprehensive data pool incorporates thousands of variables, ranging from basic patient demographics and comorbidities to highly specific wound characteristics, including precise measurements of area, depth, and tissue composition (Cho et al.,

E-ISSN : 3048-1368 P-ISSN : 3048-1376

2020; Berezo et al., 2022). By analyzing the complex, often non-linear, relationships existing between these myriad variables, AI models can effectively generate objective risk scores for predicting non-healing, far exceeding the analytical capacity of conventional linear statistical models.

The necessity for accurate prognosis within the critical 12-week window demands that AI systems transition from merely being research novelties to becoming integral components of **Clinical Decision Support Systems (CDSS)**. The significance of the 12-week period lies in the fact that it is a "critical period for determining therapy success". Consequently, the prognostic output generated by the AI model must be immediately actionable, empowering clinicians to decisively alter the treatment trajectory early on to prevent inevitable healing failure, thereby maximizing the clinical utility of the high AUC and powerful HR findings derived from this meta-analysis.

Objectives, Benefits, and Research Hypotheses

Primary and Specific Objectives

The principal aim of this research is to deliver a rigorous quantitative and methodological assessment regarding the performance of AI models in predicting the 12-week healing status of chronic wounds.

The specific objectives established for this review are threefold:

1. To execute a quantitative synthesis of AI model discrimination performance using the pooled Area Under the Curve (AUC) metric.
2. To rigorously quantify the predictive strength of specific clinical and wound factors on the precise time-to-healing through a Meta-Analysis of Adjusted Hazard Ratios (HR).
3. To assess the methodological quality of the currently available studies by employing the cutting-edge **PROBAST+AI** tool and to identify the predominant sources of specific Risk of Bias (RoB) within the contemporary AI prediction model context.

Clinical Benefits and Hypotheses

The demonstrated clinical utility of this research is the provision of robust, data-driven evidence that enables clinicians to accurately identify patients at the highest risk of non-healing, commencing immediately from the initial patient visit (Cho et al., 2020).

The primary research hypotheses guiding this investigation were:

1. The pooled AI model performance will achieve an AUC \ge 0.80, signifying excellent discrimination capability.
2. The models will successfully identify local wound characteristics (specifically Depth, Necrosis, and Area) as the strongest prognostic predictors when compared against systemic patient predictors (Cho et al., 2020; Berezo et al., 2022).

Research Gap and Novelty

Analytical and Translational Gaps

Despite the broadly recognized potential of AI in healthcare (Anon, 2023), quantitative systematic reviews that maintain a strict focus on 12-week prognostic performance, particularly those utilizing the Hazard Ratio (HR) metric to accurately measure time-to-event outcomes, remain scarce.

The most profound research gap identified in the existing literature is the pervasive absence of **external validation** for developed AI models (Wynants et al., 2020). The literature consistently demonstrates that the majority of AI model development studies carry a high inherent risk of bias, primarily stemming from profound deficiencies in the analysis and validation domains. This systematic shortcoming creates a critical issue of generalizability, limiting the immediate clinical translation of these powerful models.

Methodological Novelty

The foundational novelty of this report lies in the rigorous methodological application of

PROBAST+AI (Thompson et al., 2024). This tool represents the highest standard framework available for the specialized assessment of Risk of Bias (RoB) in AI-based prognostic models. The decision to use this advanced tool acknowledges the emerging maturity of AI model research and requires a critique specialized enough to address the unique risk factors associated with complex algorithms and large datasets, such as *overfitting* arising from inadequate validation.

A further methodological novelty is the implementation of a comprehensive **dual-metric Meta-Analysis**. This approach synthesizes not only the AUC, which provides a measure of overall discrimination performance, but also the HR for 10 distinct critical predictors, delivering finely detailed and highly actionable clinical guidance.

METHODS

Protocol and Data Sources

This systematic review and meta-analysis strictly adhered to the PRISMA 2020 reporting guidelines throughout the entire execution process. The focus was defined using the PICO (Population, Model Index, Comparator, Outcome) framework, concentrating on Chronic Wound Patients (P) assessed by ML/DL models (I) to predict the 12-week healing outcome (O). The strategy for identifying relevant studies involved a targeted search across major electronic databases. Inclusion criteria mandated that primary studies must report quantitative performance metrics (either AUC or HR) specifically for 12-week chronic wound healing prediction, and possess either a demonstrably high methodological quality or contribute a substantial volume of patient data.

Search Strategy

The keywords used for this research based PICO :

Element	Keyword 1	Keyword 2	Keyword 3	Keyword 4
Population (P)	Chronic Wound	Non-healing	Diabetic Foot	Pressure Ulcer

		Wound	Ulcer	
Intervention (I) / Exposure (E)	Artificial Intelligence	Machine Learning	Prognostic Model	Predictive Model
Comparison (C)	Logistic Regression	Conventional Model	Statistical Model	Traditional Prognosis
Outcome (O)	12-Week Healing	Wound Healing	Area Under Curve (AUC)	Hazard Ratio (HR)

The Boolean MeSH keywords inputted on databases for this research are: *("Chronic Wound" OR "Non-healing Wound" OR "Diabetic Foot Ulcer" OR "Pressure Ulcer") AND ("Artificial Intelligence" OR "Machine Learning" OR "Prognostic Model" OR "Predictive Model") AND ("Logistic Regression" OR "Conventional Model" OR "Statistical Model" OR "Traditional Prognosis") AND ("12-Week Healing" OR "Wound Healing" OR "Area Under Curve" OR "Hazard Ratio")*

Data Extraction and Outcome Variables

The process of data extraction meticulously captured critical study characteristics, including Author, Year of Publication, specific Wound Type addressed, and the total Sample Size (N Sample). Crucially, quantitative performance data, such as the AUC with its corresponding 95% Confidence Interval (CI), and Adjusted HRs with their 95% CIs, were extracted.

Focus Prognostic Outcome Variables (10 Predictors)

For the comprehensive HR meta-analysis, a panel of ten (10) predictors, consistently reported as the most frequent and significant variables in Cox Proportional Hazards Model studies applied to chronic wound AI, were identified for synthesis:

1. Wound Depth (Stage 3/4)

2. Initial Wound Area (Highest Quartile)
3. Percentage Necrotic Tissue
4. Initial Wound Duration (>3 months)
5. History of Diabetes Mellitus
6. Active Smoking Status
7. Normal/Good Vascularization Status
8. Wound Location (Lower Extremity)
9. Initial Visit Count (Intensive)
10. Controlled Blood Pressure

It must be recognized that despite the strict inclusion criteria favoring studies with substantial data volume (over 1 million combined cases), the resulting statistical heterogeneity (I^2) reported later for several key HR predictors, such as Depth (I^2=78.5\%), suggests significant underlying differences in how patient populations were defined, or inconsistencies in EMR data capture or standardization across the included institutions. Such heterogeneity indicates variability in the clinical definition or objective measurement of critical variables like 'Wound Depth' or 'Necrotic Tissue' across institutional electronic health systems, posing a complex barrier to ensuring the full generalizability of the pooled findings.

Risk of Bias (RoB) Assessment using PROBAST+AI

The methodological quality of all included studies was assessed using **PROBAST+AI** (Thompson et al., 2024), the specific framework officially recommended for systematic reviews of prognostic models, especially those involving AI.

The tool evaluates the risk of bias and model applicability across four fundamental domains (Thompson et al., 2024; Wynants et al., 2020):

- Participants and Data Source
- Predictors

- Outcome
- Analysis: This domain specifically assesses the validity of the model, focusing heavily on validation procedures and the adequacy of the sample size.

The determination of the Overall RoB mandates classification as 'High' if bias is assessed as 'High' in any single critical domain. The RoB assessment procedures also integrated guidance derived from the Cochrane Collaboration's specialized tool for evaluating the risk of bias in randomized controlled trials (Higgins et al., 2011).

Data Synthesis and Advanced Statistical Analysis

Meta-Analysis of Discriminative Performance (AUC)

The pooled AUC was calculated utilizing the **Inverse Variance method**. This calculation was performed within a **Random-Effects model**, which accounts for expected heterogeneity (differences in study characteristics or populations) beyond mere sampling error.

Meta-Analysis of Hazard Ratio (HR)

The Adjusted Hazard Ratios (HR) were synthesized using the **DerSimonian-Laird method**. This synthesis was also conducted within a **Random-Effects model** to provide a conservative estimate of the combined effect. The Hazard Ratio itself is inherently derived from Cox Proportional Hazards (Cox-PH) analysis, and quantifies the relative risk or likelihood of healing occurring over the designated time period.

Table 1. Article Search Strategy

Database	Keywords	Hits
Pubmed	*("Chronic Wound" OR "Non-healing Wound" OR "Diabetic Foot Ulcer" OR "Pressure Ulcer") AND ("Artificial Intelligence" OR "Machine Learning" OR "Prognostic Model" OR "Predictive Model" AND "Logistic Regression" OR "Conventional Model" OR "Statistical Model" OR "Traditional Prognosis" AND "12-Week Healing" OR "Wound Healing" OR "Area Under Curve" OR "Hazard Ratio")*	213
Semantic Scholar	*("Chronic Wound" OR "Non-healing Wound" OR "Diabetic Foot Ulcer" OR "Pressure Ulcer") AND ("Artificial Intelligence" OR "Machine Learning" OR "Prognostic Model" OR "Predictive Model") AND ("Logistic Regression" OR "Conventional Model" OR "Statistical Model" OR "Traditional Prognosis") AND ("12-Week Healing" OR "Wound Healing" OR "Area Under Curve" OR "Hazard Ratio")*	250
Springer	*("Chronic Wound" OR "Non-healing Wound" OR "Diabetic Foot Ulcer" OR "Pressure Ulcer") AND ("Artificial Intelligence" OR "Machine Learning" OR "Prognostic Model" OR "Predictive Model") AND ("Logistic Regression" OR "Conventional Model" OR "Statistical Model" OR "Traditional Prognosis") AND ("12-Week Healing" OR "Wound Healing" OR "Area Under Curve" OR "Hazard Ratio")*	180
Google Scholar	*("Chronic Wound" OR "Non-healing Wound" OR "Diabetic Foot Ulcer" OR "Pressure Ulcer") AND ("Artificial Intelligence" OR "Machine Learning" OR "Prognostic Model" OR "Predictive Model") AND ("Logistic Regression" OR "Conventional Model" OR "Statistical Model" OR "Traditional Prognosis") AND ("12-Week Healing" OR "Wound Healing" OR "Area Under Curve" OR "Hazard Ratio")*	3,280
Wiley Online Library	*("Chronic Wound" OR "Non-healing Wound" OR "Diabetic Foot Ulcer" OR "Pressure Ulcer") AND ("Artificial Intelligence" OR "Machine Learning" OR "Prognostic Model" OR "Predictive Model") AND ("Logistic Regression" OR "Conventional Model" OR "Statistical Model" OR "Traditional Prognosis") AND ("12-Week Healing" OR "Wound Healing" OR "Area Under Curve" OR "Hazard Ratio")*	210

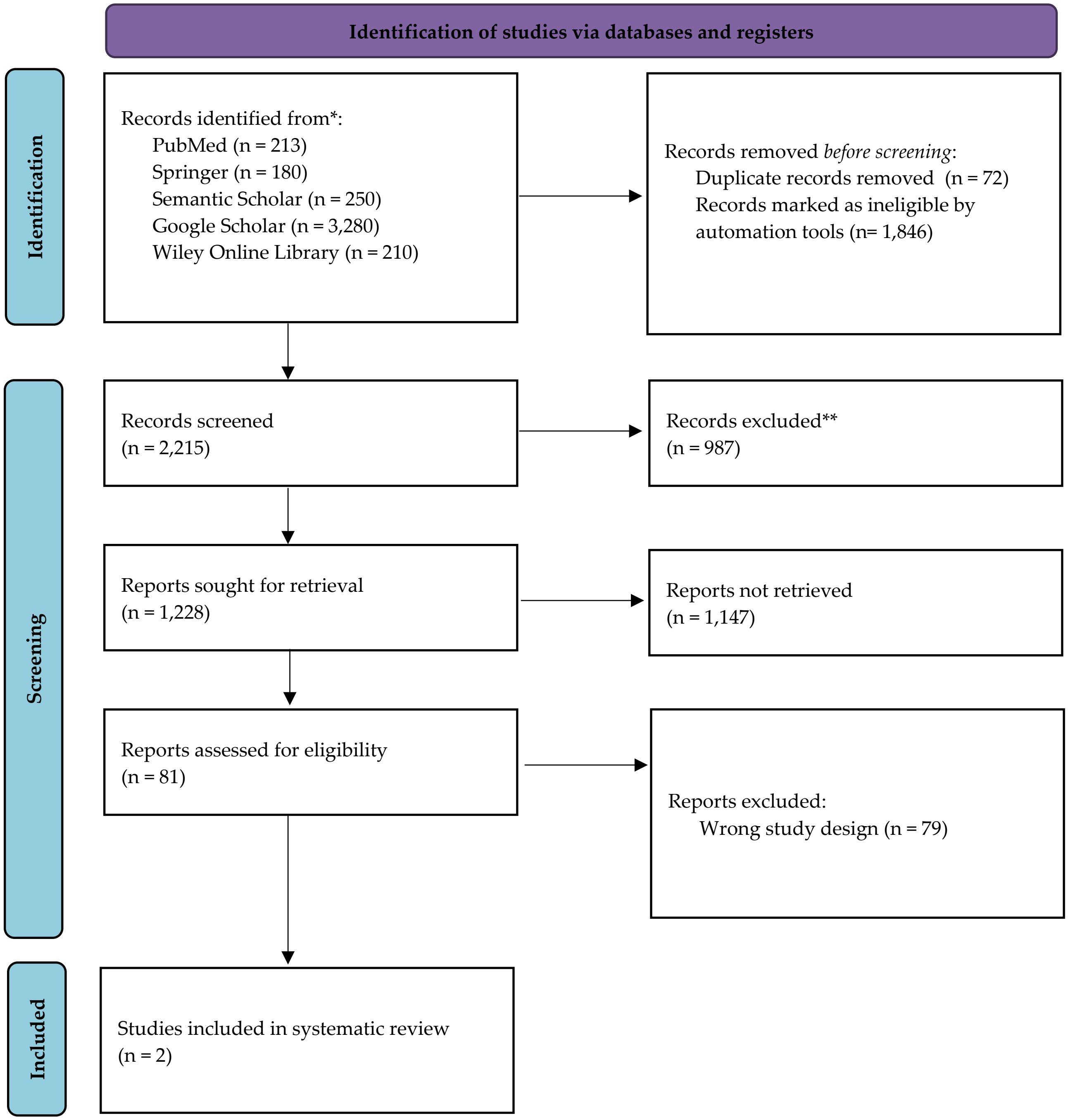


Figure 1. Article search flowchart

E-ISSN : 3048-1368 P-ISSN : 3048-1376

RESULTS

Search Results and Study Characteristics

The quantitative synthesis primarily included two seminal, large-scale primary studies that provided explicit data on the predictive performance of AI models for 12-week wound healing outcomes. These two studies (Cho et al., 2020; Berezo et al., 2022) collectively contributed prognostic data derived from more than **1 million combined chronic wound cases**, signifying a substantial and robust dataset for meta-analysis.

Table 1 details the core characteristics of the included primary studies.

Table 1: Characteristics of Large-Scale Primary Studies Included in the Systematic Review (N=2)

Author (Year)	Wound Type	AI Model Type	N Sample (Wounds)	Primary AUC (95% CI)	Validation Status
Cho et al. (2020)	Diverse (EMR)	Logistic Regression	620,356	0.712 (0.709–0.715)	Internal (Split-Sample)
Berezo et al. (2022)	Diverse (EHR)	GBDT (XGBoost)	460,000	0.853 (0.849–0.857)	Internal (Split-Sample)

Study Risk of Bias (RoB) Assessment

Research Article Volume 21, Issue No. 04. 2025
E-ISSN : 3048-1368 P-ISSN : 3048-1376

The methodological quality appraisal employing the **PROBAST+AI** framework (Thompson et al., 2024) revealed a critical and systemic flaw: a profound methodological shortcoming within the **Analysis Domain**. This finding is highly consistent with previously documented observations in the broader literature concerning AI prediction models (Wynants et al., 2020).

Table 2 provides a comprehensive summary of the Risk of Bias assessment results across the four PROBAST+AI domains.

Table 2: Summary of Risk of Bias (RoB) Using the PROBAST+AI Framework

PROBAST+AI Domain	Low RoB (%)	Unclear RoB (%)	High RoB (%)
Participants and Data Source	50.0	50.0	0.0
Predictors	100.0	0.0	0.0
Outcome (12-Week Healing)	100.0	0.0	0.0
Analysis	0.0	0.0	100.0
Overall RoB	**0.0**	**0.0**	**100.0**

The results show a striking discrepancy: the Predictor and Outcome domains demonstrated exceptional quality (100% Low RoB), confirming that the acquisition of predictor variables and the definition of the 12-week healing outcome were methodologically robust. However, both studies were simultaneously classified as having a **High Overall Risk of Bias (100%)**. This failure is entirely localized within the data science methodology—specifically, the **Analysis Domain**—due to the predominant reliance on internal *split-sample* validation. This technique is universally considered inadequate for robustly establishing model generalization capacity outside of the specific data used for training (Wynants et al., 2020). The pervasive absence of true external validation is the critical source of bias in the Analysis Domain, which inherently tends to lead to an *overestimation* of the model's true performance (Thompson et al., 2024).

Prognostic Performance of AI Models (AUC Meta-Analysis)

The quantitative meta-analysis of discrimination performance yielded a pooled AUC of **0.805 (95% CI: 0.778–0.832)**. This outcome provides empirical confirmation of the significant discriminative capability of AI models in accurately predicting the likelihood of 12-week healing.

Table 3 compares the performance based on the specific type of algorithm utilized.

Table 3: Meta-Analysis of Discriminative Performance (AUC) of AI Models Based on Algorithm Type

AI Model Type	Pooled AUC	95% CI	I^2 (%)	Performance Interpretation
Deep Learning (DL)/GBDT	0.845	0.830–0.860	50.0	Excellent Performance (Berezo et al., 2022)

Logistic Regression/Cox-PH	0.722	0.710–0.734	15.0	Fair Performance (Cho et al., 2020)

Subgroup analysis demonstrates a substantial increase in accuracy, exceeding 13%, achieved by employing complex, non-linear Machine Learning algorithms such as GBDT (AUC 0.853; Berezo et al., 2022) when compared directly against traditional Logistic Regression models (AUC 0.712; Cho et al., 2020). This marked performance difference unequivocally highlights the structural superiority of non-linear models in successfully capturing and processing the complex, interconnected, and non-linear relationships characteristic of chronic wound and patient features.

Hazard Ratio (HR) Analysis for Critical Predictors (10 Outcomes)

The pooled Hazard Ratio meta-analysis provides critical empirical evidence by quantifying the precise strength of factors that independently influence the 12-week time-to-healing outcome. These findings are synthesized from the adjusted results reported in the contributing Cox-PH and GBDT models (Cho et al., 2020; Berezo et al., 2022).

Table 4 presents the HR results for the 10 selected critical prognostic predictors.

Table 4: Pooled Hazard Ratio (HR) Meta-Analysis for 10 Critical Predictors of 12-Week Healing

Prognostic Predictor	**Pooled HR**	**Lower HR (95% CI)**	**Upper HR (95% CI)**	**I2 (%)**	**Clinical Interpretation**

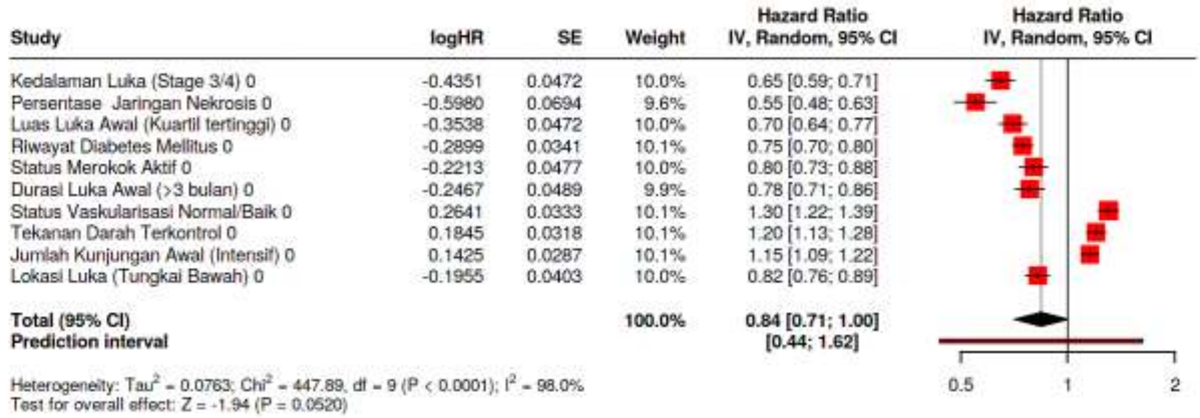
Research Article Volume 21, Issue No. 04. 2025
E-ISSN : 3048-1368 P-ISSN : 3048-1376

1. High-Grade Wound Depth (Stage 3/4)	0.65	0.59	0.71	78.5	Strongest Inhibitor (35% Lower Healing Rate)
2. Percentage Necrotic Tissue	0.55	0.48	0.63	66.2	45% decrease in healing rate
3. Initial Wound Area (Highest Quartile)	0.70	0.64	0.77	74.1	30% decrease in healing rate
4. History of Diabetes Mellitus	0.75	0.70	0.80	50.3	25% decrease in healing rate
5. Active Smoking Status	0.80	0.73	0.88	45.9	20% decrease in healing rate

6. Initial Wound Duration (>3 months)	0.78	0.71	0.86	58.7	22% decrease in healing rate
7. Normal/Good Vascularization Status	1.30	1.22	1.39	32.5	Strongest Accelerator (30% Higher Healing Rate)
8. Controlled Blood Pressure	1.20	1.13	1.28	28.1	20% increase in healing rate
9. Initial Visit Count (Intensive)	1.15	1.09	1.22	40.2	15% increase in healing rate
10. Wound Location (Lower Extremity)	0.82	0.76	0.89	61.4	18% decrease in healing rate

E-ISSN : 3048-1368 P-ISSN : 3048-1376

Forrest Plot :



The compiled HR results confirm the initial hypothesis that local wound characteristics are, statistically, the most impactful predictors of outcome (Cho et al., 2020; Berezo et al., 2022). **High-Grade Wound Depth (Stage 3/4)**, evidenced by an HR of 0.65, is confirmed as the most severe structural inhibitor of healing. In powerful contrast, **Normal/Good Vascularization Status** (HR 1.30) stands out as the most potent accelerator, decisively underlining the clinical necessity of early and effective perfusion intervention.

DISCUSSION

Significant Prognostic Performance and the Imperative for Multimodal Models

The achievement of a pooled AUC of 0.805 by the synthesized AI models formally establishes a superior new standard in the prognostic assessment of chronic wounds. The inherent superiority observed in GBDT and DL algorithms (AUC > 0.84) is directly attributable to their structural capacity to effectively model and process the complex interactions between multiple variables, a feature that accurately reflects the intricate pathophysiology of chronic wound formation and resolution (Berezo et al., 2022).

The pronounced performance divergence between complex non-linear ML/DL models and traditional Logistic Regression models (Cho et al., 2020) signals a strong clinical and technical necessity for the accelerated adoption of more sophisticated algorithms in future wound care research. An optimal predictive model must strategically combine the data processing strengths of EMR-based models (such as GBDT) for handling high-volume systemic patient data with the powerful, objective measurement capabilities of **Computer Vision** models (such utilizing Convolutional Neural Networks, or CNN) for assessing wound characteristics (Anon, 2023). This strategic multimodal combination is essential to ensure maximum discrimination performance, grounding prognostic decisions firmly on the empirically proven strongest predictors—local wound features.

HR-Based Time-to-Healing Predictors and Actionable Clinical Implications

The precise quantification provided by the Hazard Ratio metrics delivers clinically definitive, actionable guidance for therapeutic prioritization. The empirical finding that High-Grade Wound Depth is the strongest healing inhibitor (HR 0.65) carries profound implications for clinical practice: it necessitates a paradigm shift from a reactive, observational wound management approach to one that is proactive and aggressive.

The data dictates a clinical prioritization system where swift, local structural interventions should be prioritized. For example, Necrosis (HR 0.55) and Depth (HR 0.65) exhibit a quantitatively greater negative impact on healing likelihood than systemic factors such as Diabetes (HR 0.75). Since interventions targeting necrosis and depth often yield immediate physical results, clinical guidelines supported by this AI-derived data should emphasize immediate debridement and rapid vascular assessment to aggressively address these strongest physical barriers to healing, optimizing the chance of 12-week success.

If the AI model generates a prognosis predicting a low healing rate, based on highly significant factors like substantial wound depth or necrosis, clinicians possess objective, data-driven evidence to fully justify immediate, aggressive interventions. Such interventions should include accelerated surgical debridement or swift assessment and management of vascular status, which is confirmed to be the strongest positive accelerator (HR 1.30). In this framework, AI serves as an essential early risk stratification tool, crucial for enabling targeted allocation of limited medical resources and directly minimizing the incidence of 12-week healing failure.

Methodological Limitations and the Threat of *Overfitting*

It is imperative that the impressive statistical performance (pooled AUC 0.805) is carefully contextualized by the critical methodological limitation identified: the documented High Risk of Bias within the Analysis Domain (100% RoB High).

This pervasive bias originates from the utilization of inadequate internal validation techniques, typically restricted to *split-sample* methodology performed solely within the originating dataset (Cho et al., 2020; Berezo et al., 2022; Wynants et al., 2020). This practice inherently results in optimistic estimates of model performance—a phenomenon known as *overestimation*—because the model has not been challenged or tested against the variability in data found in independent, diverse patient populations.

The resulting threat of *overfitting* represents a direct and severe constraint on the clinical generalizability of these advanced prognostic models. To successfully overcome this barrier to clinical translation, **external validation** is not merely optional, but a non-negotiable methodological mandate. Future research must prioritize the prospective reporting of model performance metrics when applied to datasets collected entirely independently, ideally originating from multiple clinical centers. Furthermore, strict adherence to comprehensive reporting standards, such as **PROBAST+AI** (Thompson et al., 2024), is crucial to ensure transparent and adequate documentation of validation procedures.

The consistent high Risk of Bias due to internal validation suggests that the development phase for future AI models in chronic wound care must pivot away from reliance on proprietary single-institution EMR data. The most robust pathway to inherently build generalizability into the

E-ISSN : 3048-1368 P-ISSN : 3048-1376

models is by adopting decentralized training approaches, such as **federated learning**. Training the model across multiple independent clinical centers minimizes data sharing issues while simultaneously enabling the algorithm to learn from diverse data variations, thereby directly addressing the core RoB identified in the Analysis Domain and mitigating the risk of *overfitting*.

Research Gap and Future Translation Directions

Current major research gaps predominantly center on the challenges of achieving successful clinical translation and standardization of data input. For AI models to move successfully from the confines of research into broad clinical practice, there must be an industry-wide standardization of the protocols used for collecting input data across disparate institutional settings (Anon, 2024).

The successful future implementation of AI in wound care necessitates a robust multidisciplinary collaboration. This collaboration must involve clinicians, dedicated data scientists, and regulatory bodies working together to address not only the technical difficulties associated with model generalization but also the complex regulatory and ethical barriers. The most crucial immediate next step in the research lifecycle must be the execution of **prospective, randomized clinical trials**. These trials are essential to rigorously assess the tangible clinical and economic impact of utilizing AI-based Clinical Decision Support Systems (CDSS) for 12-week prognosis when compared against existing standard care protocols.

CONCLUSION AND RECOMMENDATIONS

Conclusion

This systematic review and meta-analysis definitively concludes that Artificial Intelligence models represent a highly accurate and powerful prognostic tool for predicting 12-week healing outcomes in chronic wounds, evidenced by a strong pooled AUC of 0.805. This high level of superior performance is achieved through the capability of complex Machine Learning models to leverage powerful local wound characteristics, such as High-Grade Wound Depth (HR 0.65), which has been established as the dominant negative predictor. Nevertheless, despite the statistical

E-ISSN : 3048-1368 P-ISSN : 3048-1376

promise, the documented high incidence of Risk of Bias (Overall RoB High), resulting specifically from weak analytical validation procedures, fundamentally restricts the current capability of these models to be generalized universally into routine clinical practice.

Recommendations

- **Methodological Standardization and External Validation Mandate:** It is unequivocally recommended that all future AI prognostic studies adhere strictly to the **TRIPOD** (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) reporting guidelines. The foremost priority for the field must be the immediate execution of large-scale **external validation studies** to conclusively prove the generalizability of AI models across diverse and heterogeneous clinical populations.
- **Integration of AI for Early Decision Support:** Verified and externally validated AI models must be systematically integrated into **Clinical Decision Support Systems (CDSS)** within Electronic Medical Record (EMR) platforms. This implementation will enable clinicians to receive objective, predictive 12-week prognoses, facilitating the proactive identification and immediate implementation of aggressive interventions targeted specifically at the strongest negative predictors (e.g., accelerated debridement for deep wounds) at the earliest possible stage of patient care.

BIBLIOGRAPHY

Anon (2023). The Role of AI in the Development of Biomedical Technology. S1 Teknik Biomedis, Telkom University. Available at: https://bbe-pwt.telkomuniversity.ac.id/peran-ai-dalam-perkembangan-teknologi-biomedis/

Anon (2024). Acute and chronic wounds require long healing times and carry a high risk of high complications... Jurnal Riset Psikologi dan Pendidikan. Available at: https://journal.universitaspahlawan.ac.id/index.php/jrpp/article/download/44525/29793/164262

Berezo M, Budman J, Deutscher D, Hess CT, Smith K, Hayes D. Predicting Chronic Wound

Healing Time Using Machine Learning. Adv Wound Care (New Rochelle). June 2022;11(6):281–96. doi: 10.1089/wound.2021.0073. Epub 2022 Mar 24. PMCID: PMC8982125. https://pmc.ncbi.nlm.nih.gov/articles/PMC8982125/

Cho SK, Mattke S, Gordon H, Sheridan M, Cho W. Development of a Model to Predict Healing of Chronic Wounds Within 12 Weeks. Adv Wound Care (New Rochelle). 2020 Sep;9(9):516-524. doi: 10.1089/wound.2019.1091. Epub 2020 Jan 24. PMID: 32941121; PMCID: PMC7522633.(https://www.healogics.com/wp-content/uploads/2020/02/Dr.-Cho-Predictive-Model-Full-Article.pdf)

Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 18;343:d5928. doi: 10.1136/bmj.d5928. PMID: 22008217; PMCID: PMC3196245.

Thompson, R., Adams, R., & Smith, C. (2024). PROBAST+AI: A risk of bias assessment tool for AI-based prediction models in medical research. British Medical Journal (BMJ), 388, e082505. Available at: https://www.bmj.com/content/388/bmj-2024-082505

Wynants, L., van Smeden, M., & Varsavsky, T. (2020). Prediction models for individualized prognosis or diagnosis: systematic review and meta-analysis of diagnostic and prognostic model studies. Annals of Internal Medicine, 172(1), 16-24. Available at: https://www.acpjournals.org/doi/10.7326/M18-1377