Towards an AI-based Objective Prognostic Model for Quantifying Wound Healing

Rishabh Gupta, Lucas Goldstone, Shira Eisen, Dhanesh Ramachandram, Amy Cassata, Robert D. J. Fraser, Jose L. Ramirez-GarciaLuna, Robert Bartlett and Justin Allport

Abstract— **Chronic wounds affect millions of people worldwide every year. An adequate assessment of a wound's prognosis is a critical aspect of wound care since it assists clinicians in understanding wound healing status, severity, triaging and determining the efficacy of a treatment regimen, thus guiding the clinical decision making. The current standard of care involves using wound assessment tools, such as Pressure Ulcer Scale for Healing (PUSH) and Bates-Jensen Wound Assessment Tool (BWAT), to determine wound prognosis. However, these tools involve manual assessment of a multitude of wound characteristics and skilled consideration of a variety of factors, thus, making wound prognosis a slow process which is prone to misinterpretation and high degree of variability. Therefore, in this work we have explored the viability of replacing subjective clinical information with deep learning-based objective features derived from wound images, pertaining to wound area and tissue amounts. These objective features were used to train prognostic models, that quantified the risk of delayed wound healing, using a dataset consisting of 2.1 million wound evaluations derived from more than 200,000 wounds. The objective model, which was trained exclusively using image-based objective features, achieved at minimum a 5% and 9% improvement over PUSH and BWAT, respectively. Our best performing model, that used both subjective and objective features, achieved at minimum an 8% and 13% improvement over PUSH and BWAT, respectively. Moreover, the reported models consistently outperformed the standard tools across various clinical settings, wound etiologies, sexes, age groups and wound ages, thus establishing the generalizability of the models.**

Index Terms— **Chronic wounds, prognosis, health monitoring, contactless monitoring, wound imaging**

I. INTRODUCTION

Chronic wounds impact a large segment of the world's population, and are estimated to affect 1-2% of the population in the developed nations [1]. These wounds fail to heal in an orderly and timely manner [2], imposing a considerable financial burden on the healthcare systems of these countries and negatively impacting the quality of life

R. Gupta, L. Goldstone, S. Eisen, D. Ramachandram, A. Cassata, R. D. J. Fraser, J. L. Ramirez-GarciaLuna, R. Bartlett and J. Allport are with the Swift Medical Inc., Toronto, ON M5H 3W4, Canada (e-mail usernames: rishabh.gupta, lucas.goldstone, shira.eisen, dhanesh.ramachandram, amy.cassata, rob.fraser, jose.luna, bob.bartlett, justin.allport; domain: swiftmedical.com).

of patients suffering from chronic wounds [3]. Within the USA, an estimated $25 billion is spent every year on wound management and treatment costs [1]. Despite this massive burden, efficient wound care treatment remains a challenge even for experienced clinicians. Clinicians rely on standard prognostic tools based on subjective wound assessments to identify and triage worsening wounds, as well as to provide reliable and quantifiable wound outcomes (e.g., probability of delayed healing or amputation) to measure the efficacy of different clinical interventions [4].

A typical wound assessment entails examining *wound extent*, *burden* and *severity* [2]. The *wound extent* is determined by measuring the wound dimensions, e.g., area and depth, and amounts of different tissues, e.g., epithelial and necrotic, present within the wound bed. The amount of necrotic and slough tissue within the wound bed has been found to be directly related to the worsening of wounds [5]. The *wound burden* is a function of *wound extent* and other attributes, such as infection, inflammation, and wound edges. The amount and type of exudate, along with markers of inflammation such as induration (firmness of tissues with margin) and edema (shiny and taut skin), help care providers flag local wound infection[6]. On the other hand, the condition of wound edges and surrounding tissue is an effective indicator of wound healing. For example, the presence of attached wound edge along with an advancing border of epithelium can indicate that a wound is healing, whereas the presence of rolled edges can indicate that a wound's healing progress has stalled [7]. The *wound severity* is guided by wound burden, patient, and environmental factors, such as age, sex, ethnicity, socio-economic status, co-morbidities (e.g., diabetes mellitus, renal disease) and systemic agents (e.g., dialysis, vasoactive drugs). Together, these attributes can influence the access to as well as the quality of care needed for ideal wound healing [2].

Standard wound assessment tools used in clinical practice focus on subjectively quantifying *wound extent* and *burden*. A majority of the wound assessment tools have been developed for pressure injuries specifically. For instance, the Pressure Ulcer Scale for Healing (PUSH) [8], specifically tracks the healing progress of pressure injuries. PUSH uses only three wound characteristics, which are wound area, exudate amount and presence of a particular tissue within the wound bed. The PUSH features were found to account for only 39-57% of the variation in wound progress over time [9], [10]. Therefore, other tools, such as, DESIGN [11], Wound Healing Scale [12],

Sessing Scale [13] and Sussman Wound Healing Tool [14], attempted to develop an improved tool to monitor the healing of pressure injuries. These tools additionally rely upon characterizing wound depth, infection, maceration, hemorrhage, location and wound edges. Similarly, prognostic tools, such as PEDIS [15], SINBAD [16] and diabetic foot ulcer assessment scale [17], have been developed specifically for monitoring diabetic wounds. The validity of these tools across different wound types is limited [18]. On the other hand, the Bates-Jensen Wound Assessment Tool (BWAT) is a prognostic tool that has been used in various clinical trials and clinical settings across chronic wound from a variety of etiologies [19]. Using BWAT, clinicians monitor wound healing by subjectively assessing eleven wound attributes, including wound area, depth, edges, tissue types, exudate amounts and types, skin colour, induration and edema.

The subjective assessment of wound characteristics is, time consuming, susceptible to misinterpretation and high inter-rater variability [20], [21]. Moreover, using integer numbers for subjectively scoring wound characteristics leads to lower sensitivity in tracking progress of smaller wounds or wounds in the final stages of healing in comparison to larger wounds [9]. Furthermore, the clinicians' level of expertise also contributes to the accuracy of the assessment [22]. The camera-based monitoring of wounds offers a non-contact way of assessing wounds, while improving the sensitivity, accuracy and speed of wound assessments [23], [24]. To this end, this study explores the feasibility of replacing subjective factors, related to *wound extent*, with their objective representations for wound prognosis, derived from wound images obtained using Swift Medical's wound imaging platform. Swift Medical provides a wound imaging platform that has been demonstrated to be quicker [24], reproducible and accurate [25], as well as having robust AI-integration for assessing wound areas, margins and tissue composition [21], and is therefore ideal for the development of objective metrics. Additionally, the wide adoption of the solution has enabled a large dataset, for example a recent study of Swift's real-world dataset was ten times larger than any previously published study [26]. This rich dataset is ideal for the development of machine learning models for wound care prognostics.

Deep learning based models have been found to perform accurately on wound and tissue segmentation tasks, as reported in [21], [27]–[30]. Therefore, these models are ideal candidates for objectively determining wound characteristics, such as wound size and tissue amounts, that quantify *wound extent*. However, the prognostic capability of such techniques is not very well understood. There have been some attempts to develop an end-to-end system for wound prognosis using wound size determined by wound segmentation models [31], [32]. However these models were evaluated on a limited set of wound etiologies and time-series. Moreover, these did not evaluate the prognostic capability of deep learning based tissue segmentation techniques.

Therefore, in the present study we have attempted to establish the prognostic capability of a system that replaces subjective representation of all aspects of *wound extent*, as shown in Fig 1. To replace traditional clinician measurements of wound

TABLE I: Total wound series per wound type across skilled nursing facilities (SNF) and home healthcare facilities (HHF)

Wound Type	SNF	HHF	Total
PIs	138,125	5,353	143,478 (71.1%)
Arterial	11,626	383	12,009 (6.0%)
Diabetic	16,011	2,842	18,853 (9.4%)
Venous	23,838	3,285	27,123 (13.5%)
Total	189,600 (94.1%)	11,863 (5.9%)	201,463

PIs: Pressure Injuries

extent, we used wound segmentations and tissue amounts from two pre-existing models called AutoTrace and Autotissue, as described in [21]. These models were considered ideal for the task as these were trained using significantly larger dataset and validated across a larger variety of wound etiologies compared to other deep learning models. The estimated wound area and tissue amounts, along with other wound characteristics, were then used to train prognostic models using the Cox Proportional Hazards (Cox-PH) model [33]. These models are easily interpretable and can be used for providing quick and actionable insights to the clinicians related to wound prognosis. Furthermore, we have provided a comparison of our techniques against standard clinical prognostic tools, such as PUSH and BWAT, using the dataset acquired from different clinical settings across various patient demographics, and time-to-heal. The study also provides a guide to the development of completely objective prognostic models in the future by identifying the features correlated with wound healing. To the best of our knowledge such a comprehensive characterization of wound prognostic techniques has not been reported in previous studies.

II. METHODS

A. Dataset Description

The dataset was extracted from Swift Medical Inc's anonymized wound care database [21]. The study was deemed exempt from informed consent requirements after review from Research Review Board Inc., an independent review board located in Ontario, Canada [1]. This proprietary dataset was derived from 2,361 skilled nursing facilities (SNFs) and 141 home healthcare facilities (HHFs), spread across North America. The dataset focused on four frequently encountered wound types including pressure injuries, venous ulcers, diabetic wounds and arterial ulcers. We included wounds having at least four evaluations and were tracked for longer than three weeks, which filtered out recently acquired wounds and reduced artificially elevated levels of censoring in the dataset.

In total, the dataset consisted of 2,151,185 wound evaluations and images derived from 201,463 wounds and 98,407 patients distributed across all of the aforementioned facilities. The distribution of wound etiologies is shown in Table I, and the distribution of patient sex and age are is shown in Table II and Fig 2, respectively.

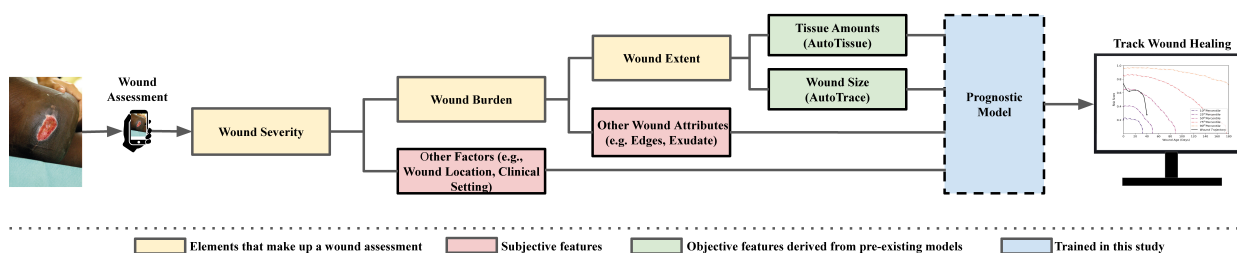[1] Study ID: Swift Medical SW-13, Research Review Board ID: (2022.W006)

Fig. 1: Individual modules of the end-to-end prognostic model trained using both subjective and objective features. Wound severity is assessed using wound, patient and environment related factors [2]. In this system, we have used pre-existing deep learning modules to quantify wound size and tissue amounts [21].

TABLE II: Sex distribution of patients across skilled nursing facilities (SNF) and home healthcare facilities (HHF)

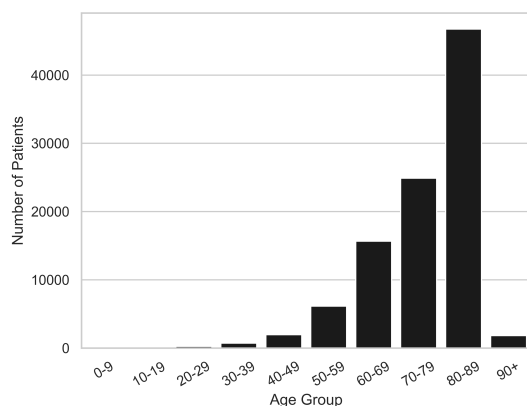| Sex	SNF	HHF	Total
Male	40,638	3,313	43,951 (44.5%)
Female	50,984	2,900	53,884 (55.0%)
Unknown	14	558	572 (0.5%)
Total	91,636 (93.0%)	6,771 (7.0%)	98,407

Fig. 2: The distribution of patients according to age group in the dataset.

B. Feature Engineering

Swift's Skin and Wound App is used across different clinical settings that include SNFs and home healthcare facilities. It was observed from our dataset that different clinical facilities leverage different wound assessment tools to monitor wound healing. This necessitates engineering features compatible with wound assessments used across facilities. In the following subsections, the subjective and objective feature engineering modules of our system are described. In the subjective feature engineering module, we have described the transformations that were used to create compatible features derived from the subjective assessment of the wound. Whereas, machine learning models and computational scaling techniques used to extract objective information, from wound images collected as part of every wound evaluation, are described in the objective feature engineering module.

1) Subjective Feature Engineering Module: Various encoding techniques were applied to wound and patient characteristics, such as integer, boolean, float and one-hot encoding for use in our new wound prognostic models (Table III). Moreover, missing values for each feature were encoded as boolean within a 'Feature Unknown' category [2], and wound locations were mapped to six different body locations as shown in Figure 3. The generated features were used to train new wound prognostic models. For comparison, PUSH and BWAT scores were computed based on their respective subjective assessment guidelines. Missing PUSH or BWAT scores within evaluations were imputed using a forward filling methodology, meaning that missing values within a series of evaluations were replaced with the immediately present preceding value.

2) Objective Feature Engineering Module: Objective features were extracted directly from the images taken during wound evaluations. Deep learning-based models were employed for objective determination of wound region and quantification of different tissue types within the wound region. The segmented wound regions and tissue regions were subsequently used to objectively determine the wound area and relative amounts of different tissue types within the wound bed. The module adopted for wound region segmentation and wound tissue segmentation are summarized in the following subsections, and the methods and performance are described in-depth in a separate article [21].

The wound segmentation module, called AutoTrace model, is based on a deep convolutional encoder-decoder neural network architecture with attention gates in the skip connections and several other customizations which make it suitable to run on mobile devices. The encoder block allows the model to extract features, whereas the decoder block produces a wound segmentation mask from the learned features. In total, the AutoTrace model consists of approximately 3.5 million parameters. Some specific customization within the AutoTrace model architecture included replacing normal convolution blocks with depth-wise separable convolutional layers to reduce the computations, and implementing strided depth-wise convolutions that can learn to downsample activations.

AutoTrace was trained using more than 400,000 image-label pairs with wound region labels determined by clinicians. The model was tested on 2,000 image-label pairs of various wound

---

[2]Feature Unknown was set to True when the feature values were missing and it was set to False when feature values were documented.

TABLE III: Subjective feature list and feature engineering transformations

Feature	Characteristics	Encoding	Mapped Feature
Tissue Type	Epithelial Tissue Present	Integer	1
	Granulation Tissue Present	Integer	2
	Slough Tissue Present	Integer	3
	Eschar Tissue Present	Integer	4
Exudate Amount	None	Integer	1
	Scant, light or small	Integer	2
	Moderate	Integer	3
	Heavy	Integer	4
Exudate Type	None	Integer	1
	Serous	Integer	2
	Sanguinous	Integer	3
	Serosanguinous	Integer	4
	Purulent	Integer	5
	Seropurulent	Integer	6
Edges	Epithelialization, Indistinct	One-hot	True/False
	Attached, Distinct	One-hot	True/False
	Non Attached, Well defined	One-hot	True/False
	Rolled Edge	One-hot	True/False
	Fibrotic	One-hot	True/False
Edema	None	Integer	0
	Under 4 cm	Integer	1
	Over 4 cm	Integer	2
	Under 4 cm with pitting	Integer	3
	Over 4 cm with pitting	Integer	4
Induration	None	Integer	0
	Under 2 cm	Integer	1
	2 to 4 cm, extends $< 50\%$ around wound	Integer	2
	2 to 4 cm, extends $> 50\%$ around wound	Integer	3
	Over 4 cm	Integer	4
Wound Area	Area computed from wound images	Float	0 - inf cm^2
Normalized Area	Relative change in area compared to first evaluation	Float	0 - inf
Log of Normalized Area	Log-transformed Normalized Area	Float	-inf - 0
Wound Location	See Figure 3 for body locations	One-hot	-
Wound Type	Pressure Injury	One-hot	True/False
	Diabetic Wound	One-hot	True/False
	Arterial Wound	One-hot	True/False
	Venous Ulcer	One-hot	True/False
Clinical Setting	SNFs	One-hot	True/False
	Home Health	One-hot	True/False
Age Groups	10 year buckets, See Figure 2 for age buckets	Integer	-
Sex	Male	One-hot	True/False
	Female	One-hot	True/False

types, including pressure injury, venous, diabetic and arterial wounds. The model performance was characterized using mean intersection over union (mIOU) between user traced wound regions mask (target mask) and the AutoTrace predicted wound region mask (predicted mask), which represents the ratio between the number of common pixels between the target and predicted masks to the total number of pixels present across both masks. The trained AutoTrace model achieved a mIOU of 0.86 for wound region segmentation. Segmented wound regions were then used to compute wound areas by scaling wound regions with respect to a calibrant sticker of known size, called HealX, which is typically placed on the same plane close to the wound. The computed wound area was then appended to the original wound series data, and normalized relative to the initial wound area followed by log-transformation for statistical normality [34].

The wound tissue segmentation module, called AutoTissue model, is also based on an encoder-decoder neural network architecture. However, it uses an EfficientNetB0 architecture [35] as the encoder, whereas the decoder is made up of 4 blocks, each consisting of a single 2-D bilinear upsampling layer followed by 2 depth-wise convolutional layers. In total, the AutoTissue model consists of approximately 3.8 million parameters.

AutoTissue was trained using 17,000 anonymized wound images labeled by trained labelers and curated by wound clinicians. The model segments detected wound region into 5 separate tissue categories including epithelial, granulation, slough, eschar and other, the latter of which included mostly healthy tissue and the HealX calibrant sticker. The model was
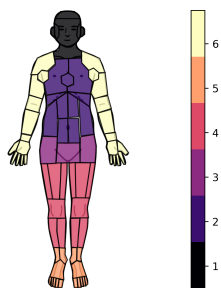
(a) Anterior Body Locations
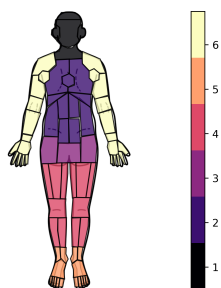
(b) Posterior Body Locations

Fig. 3: Wound locations are mapped across six color coded body locations.

tested on a set of 383 images corresponding to stage 2 pressure injuries, arterial, venous and diabetic wounds. The mIOUs for other, epithelial, granulation, slough and eschar tissue were 1, 0.42, 0.69, 0.69 and 0.85, respectively.

Wound tissue amounts were extracted for each image by first applying the AutoTrace model to focus on the wound bed, followed by the application of the AutoTissue model to determine various tissue regions within the wound bed. Next, the number of pixels within each segment was counted and the percent area of each segment was computed with respect to the total number pixels that covered the wound bed. Finally, the percentage area of each segment was log-transformed.

3) Feature Computation: As the image-based objective features were not readily available for the majority of the retrospective data collected using Swift Skin and Wound App, these models were retrospectively applied on approximately 2.1 million images collected over time. In order to reduce the processing time, the computation was executed on an Amazon Web Services (AWS) r5a.2xlarge instance-based auto-scaling cluster. Parallel batch jobs were orchestrated using Argo Workflows, an open source container-native workflow engine. Furthermore, computing objective features on the cloud mitigated the risk of raising Protected Health Information leakage concerns from wound images, as the images were securely stored on AWS and processed inside Docker containers. The computation of objective features for 2.1 million images was completed within 40 hours using 20 parallel running jobs, effectively reducing computation time by 20x compared to

running the objective feature extractor sequentially on a set of 2.1 million images using an instance based on a single central processing unit.

C. Dataset Preparation

The dataset from SNFs was stratified into training and testing sets using an 80%-20% split, resulting in 151,680 training and 37,920 testing wounds. The train/test split was performed such that the percentage of each wound type was consistent between training and testing sets. The dataset from home healthcare facilities was stratified such that the data from facilities that employ the BWAT was set aside for testing in order to have sufficient wound series for reliable testing performance. This resulted in 10,118 training and 1,745 testing wounds. Finally, the training datasets from SNFs and home healthcare facilities were combined to form a single training set. The testing sets were split into ten and three non-overlapping sub-samples for SNFs and home healthcare facilities, respectively.

D. Prognostic Model Development

Survival analysis techniques allow researchers to investigate the relationship between one or more predictor variable and the time when an event (e.g. failure of a mechanical system, death in biological organisms) occurs. Cox Proportional Hazards models estimate the effect of a predictor variable on the hazard function for the event of interest [33], and are one of the most commonly used regression techniques in survival analysis techniques. The hazard function describes the probability (or risk) of occurrence of an event based on covariate levels. The hazard function for the Cox PH model has the form:

$$h(t) = h_0(t).e^{(b_1.X_1 + b_2.X_2 + ... + b_n.X_n)}, \quad (1)$$

where, $h(t)$ is the expected hazard at time t, $h_0(t)$ is the baseline hazard which represents the hazard when all the covariates $(X_1, X_2, ...X_n)$ are set to zero and $b_1, b_2, ...b_n$ are the model weights that quantify the association between covariates and the hazard function. The model weights are computed by maximizing the partial likelihood which enables the modelling of the effects of the covariates without the need to model the hazards over time.

Commonly, the term $exp(b_n)$, is referred to as the Hazard Ratio (HR) and is used to measure the magnitude of the treatment difference [36], [37]. For example, when comparing the effect of an investigational treatment, X represents the treatment indicator. It is assigned a value of 1 for patients who receive the treatment, whereas for patients in the control group, it is assigned a value of 0. Then,

$$HR = \frac{h_0(t).exp(b \times 1)}{h_0(t).exp(b \times 0)} = exp(b), \quad (2)$$

represents the HR between the two groups. A $HR = 1$ ($b = 0$) has no effect on event probability. However, if $b > 0$, then $HR > 1$, which indicates an increase in event hazard and thus a decrease in the length of survival. If $b < 0$, then $HR < 1$, which indicates a decrease in event hazard and an increase in the length of survival.

Due to the straight-forward interpretation of model weights that is possible with Cox PH models, these models can be trained to characterize the effects of covariates on various wound outcomes (or endpoints), including a healed or infected status of a wound, or an amputation status of a limb. The resulting predictions from the trained model can be used as a prognostic index to alert clinicians of the risk of occurrence of adverse wound outcomes. In this research, a prognostic model was trained to characterize the effect of different covariates on the 'healed' or 'closed' status of the wound. Specifically, we trained our model using a variation of the Cox model called Cox time-varying proportional hazards [38] which allowed us to account for changes in covariates over time, such as increases or decreases in wound area and tissue amounts.

1) Feature Selection: We selected the features used in our models using a stepwise forward-backward selection process. In the forward step, we trained multiple univariate models and kept only the features that were significant using likelihood ratio tests. Next, in the backward selection step, various multivariate models were created by removing one feature at a time and tested against a multivariate model with all the forward selected features. The selection criteria was based on a p-value threshold of 0.05 in both steps. The lifelines package v.0.21 [39] for Python was used to train the Cox models and compute likelihood ratio tests for feature selection.

The predictive information for a subset of features was quantified using the adequacy index [40]:

$$A = LR^s/LR, \qquad (3)$$

where, LR^s is log likelihood explained by the subset of features and LR is the log likelihood explained by the entire set of features. The feature importance was computed as $1-A$, which represents the fraction of new information contributed by the feature to the model. Finally, the features with feature importance greater than a pre-defined threshold were selected.

2) Model Characterization: The selected features were used to train six different models using different combinations of features as listed in Table IV. $Model_{1,2}$ were completely based on subjectively assessed features. Specifically, $Model_1$ included features that encapsulate *wound extent* along with other *wound severity* related factors, whereas, $Model_2$ only included features encapsulating *wound extent*. In $Model_3$, we computed tissue amounts using the AutoTissue segmentation model described before, thus making $Model_3$ a hybrid model that uses both subjective and objective information. On the other hand, $Model_4$ completely replaced subjectively assessed wound area and tissue amounts with objectively computed features using AutoTrace and AutoTissue models. $Model_5$ leveraged hybrid information related to *wound extent* and *severity*, where tissue amounts were computed using the AutoTissue model. Finally, in $Model_6$ we used hybrid information that included objectively determined *wound extent* based features, and subjectively determined *wound severity* related factors.

The models were trained using the data from four different wound etiologies that included pressure injuries, diabetic, venous, and arterial wounds, thus resulting in wound-agnostic models. The trained models were compared against each other and standard wound assessment tools, such as PUSH and BWAT. Moreover, the wound-agnostic model performances were characterized for different clinical settings (i.e., SNF vs. Home-Health), sexes (i.e., Male vs. Female) and age groups (e.g., 30-39 vs. 50-59). Finally, wound-specific models were trained using the data from individual wound etiologies and their performance metrics were compared against the wound-agnostic models.

3) Performance Evaluation: Typically, the goodness-of-fit of a risk-score producing models, such as the Cox model, is quantified using the so-called Harell's Concordance Index (or C-Index) [41]. Given a prognostic model that produces risk scores 'η_i' and 'η_j' for every pair of wounds 'i' and 'j' (with $i \neq j$) with time-to-heal 'T_i' and 'T_j', respectively, a pair '(i,j)' is considered 'concordant' if '$\eta_i > \eta_j$' and '$T_i > T_j$', or '$\eta_i < \eta_j$' and '$T_i < T_j$'. However, a pair '(i,j)' is considered 'discordant' if '$\eta_i > \eta_j$' and '$T_i < T_j$', or '$\eta_i < \eta_j$' and '$T_i > T_j$'. In other words, a wound pair is concordant if the wound whose event is more imminent is given a higher risk score than the wound whose event is more distant in time. The C-Index for such a model is computed as follows:

$$C = \frac{No.\,of\,Concordant\,Pairs}{No.\,of\,Concordant\,and\,Discordant\,Pairs} \qquad (4)$$

However, Harell's C-Index is not appropriate for dynamic prediction models that incorporate longitudinal covariate data. As an alternative, the incident/dynamic time-dependent area under the Receiver Operator Characteristic ($AUC_{I/D}$) curve is suitable for dynamic prognostic models [42], [43]. The $AUC_{I/D}$ compares the predictions of incident cases, defined as the wounds that heal at the time point at which the discriminative ability is assessed, with dynamic controls, defined as the wounds that have not yet healed. The $AUC_{I/D}$ assesses the concordance of the predictions at time points 't_k' among incident cases, i.e., wounds with time-to-heal $T = t_k$, and dynamic controls, i.e., wounds with $T > t_k$ and is defined as:

$$AUC_{I/D}(t_k) = P(X_i > X_j | T_i = t_k, T_j > t_k), \qquad (5)$$

where 'X_i' and 'X_j' are the risk scores generated for two different wounds 'i' and 'j'. The $AUC_{I/D}$ directly incorporates the effects of longitudinal covariates like wound area and tissue amounts within wound bed, and can be interpreted as the probability that a random wound that healed on a given day is given a higher score than another random wound that has not yet healed on that given day. Furthermore, a concordance summary (C_t) can be estimated from $AUC_{I/D}$ as the weighted average of the area under the time-specific ROC curves using::

$$C_t = \int_t AUC_{I/D}(t).w(t)dt, \qquad (6)$$

where, $w(t) = 2.f(t).S(t)$, $f(t)$ represents the the distribution of failure times T and $S(t)$ represents the survival time [43]. We used $AUC_{I/D}$ and C_t to quantify and compare the performance of different models.

III. RESULTS

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2023.3251901

GUPTA *et al.*: TOWARDS AN AI-BASED OBJECTIVE PROGNOSTIC MODEL FOR QUANTIFYING WOUND HEALING (AUGUST 2022)　　　　7

TABLE IV: Prognostic models and the features that were used to train them

Model Type	Information	Wound Extent Features	Other Wound Severity Features
$Model_1$	Subjective	User Traced Area, Tissue Type Present	Exudate Amount, Exudate Type, Edges, Location, Setting
$Model_2$	Subjective	User Traced Area, Tissue Type Present	-
$Model_3$	Hybrid	User Traced Area, AutoTissue Amounts	-
$Model_4$	Objective	AutoTraced Area, AutoTissue Amounts	-
$Model_5$	Hybrid	User Traced Area, AutoTissue Amounts	Exudate Amount, Exudate Type, Edges, Location, Setting
$Model_6$	Hybrid	AutoTraced Area, AutoTissue Amounts	Exudate Amount, Exudate Type, Edges, Location, Setting

A. Feature Selection

The stepwise feature selection process helped identify the significance of individual features in developing the prognostic model for wounds. The feature representing Location-2 (torso and back, as shown in Figure 3) was found to be insignificant ($p > 0.05$) and was removed from further analysis. The importance of each remaining feature was quantified using the Adequacy Index, and the resulting relative feature importances are shown in Figure 4. The features with Adequacy Index greater than the empirically determined threshold of 0.5 were finally selected.
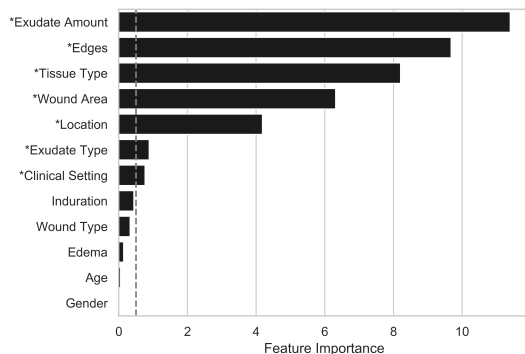

Fig. 4: Feature importance as determined using Adequacy Index. The selected features are shown with an asterisk ($*$).

B. Wound-Agnostic Model Coefficients

The selected features were used to train multiple prognostic models as outlined in previous sections. Table V lists the model coefficients and hazard ratios per feature per prognostic model. It can be noted that HRs per feature were consistently greater than or less than 1 across different models, except HRs for SNF and Location-5. Generally, other tissue amount, epithelialized and attached edges, missing features and location features had $HR > 1$, whereas the remaining features, such as wound area, slough and eschar amounts, had $HR < 1$.

C. Wound-Agnostic Model Performance

The performance for each prognostic model was determined using the concordance index (C_t) of wound pairs in the testing set. Table VI shows the C_t for each model at skilled nursing facilities and home health facilities, respectively. Moreover, it

compares the performance of each individual model across different wound types[3]. Using an Analysis of Variance (ANOVA) model, it was determined that there was a statistically significant effect of model types on C_t ($F_{9,50} = 673.52$, $p < 0.001$). Tukey's post-hoc tests revealed significant differences between all combinations except: $Model_1$ vs. $Model_3$, $Model_1$ vs. $Model_4$, $Model_3$ vs. $Model_4$ and $Model_5$ vs. $Model_6$. It can be seen that $Model_5$ and $Model_6$ were the best performing models across different clinical settings and wound types, and $Model_2$ performed the worst.

An additional ANOVA model determined that there was a statistically significant effect of wound type on model performance with $F_{4,15} > 15$, $p < 0.001$ for $Model_{3-6}$. Subsequent Tukey's post-hoc tests revealed that $Model_{3-6}$ performed statistically significantly better for pressure injuries compared to venous and diabetic wounds within SNFs. Furthermore, using Bonferroni corrected independent t-tests it was found that all prognostic models, except $Model_2$, performed significantly better ($t > 5.74$, $p < 0.0084$) under the home health setting.

Figure 5 compares the performance of prognostic models against PUSH scores at skilled nursing facilities, specifically for pressure injuries (as PUSH is not valid for other wound types). The best performing prognostic models ($Model_{5,6}$) performed at least 8-9% better than PUSH and 7% better than PUSH when missing PUSH values were imputed using a forward filling technique. ANOVA showed that there was a significant effect of model type on C_t for pressure injuries ($F_{9,70} = 673.52$, $p < 0.001$). Tukey's post-hoc tests revealed significant differences between all prognostic models and PUSH. The forward filled PUSH scores had significantly higher C_t compared to normal PUSH scores.

Figure 6 compares the performance of prognostic models against BWAT scores at home healthcare facilities for pressure injuries, venous, arterial and diabetic wounds. The best performing prognostic models ($Model_{5,6}$) performed at least 13-14% better than BWAT and 11-12% better than forward filled BWAT. ANOVA showed that there was a significant effect of model type on C_t with $F_{2,21} = 34.93$, $p < 0.001$. Tukey's post-hoc tests revealed significant differences between BWAT and all prognostic models, except $Model_2$. The forward filled BWAT score had slightly higher C_t compared to normal BWAT score, but this observation was not statistically significant.

---

[3]It should be noted that the $AUC_{I/D}$ for arterial wounds within home health setting was not reliable due to limited amount of data.

TABLE V: Model coefficients (b_n) and hazard ratios (HR), shown inside brackets, per feature for each prognostic model, along with their effect on wound healing.

Feature	$Model_1$	$Model_2$	$Model_3$	$Model_4$	$Model_5$	$Model_6$	*Effect on healing
Tissue Type Presence	-0.12 (0.89)	-0.16 (0.85)	-	-	-	-	-ve
Tissue Type Unknown	-0.06 (0.94)	-0.02 (0.98)	-	-	-	-	-ve
Log Granulation Amount	-	-	-0.02 (0.98)	-0.02 (0.98)	-0.01 (0.99)	-0.01 (0.99)	-ve
Log Slough Amount	-	-	-0.07 (0.93)	-0.07 (0.93)	-0.06 (0.94)	-0.06 (0.94)	-ve
Log Eschar Amount	-	-	-0.05 (0.95)	-0.05 (0.95)	-0.06 (0.94)	-0.06 (0.94)	-ve
Log Other Amount	-	-	0.13 (1.14)	0.13 (1.14)	0.10 (1.11)	0.10 (1.11)	+ve
Log Normalized User-traced Area	-0.07 (0.93)	-0.09 (0.91)	-0.05 (0.95)	-	-0.05 (0.95)	-	-ve
Log Normalized Auto-traced Area	-	-	-	-0.03 (0.97)	-	-0.02 (0.98)	-ve
Edges Epithelialization	0.17 (1.19)	-	-	-	0.13 (1.14)	0.15 (1.16)	+ve
Edges Attached	0.13 (1.14)	-	-	-	0.14 (1.15)	0.13 (1.14)	+ve
Edges Non Attached	-0.56 (0.57)	-	-	-	-0.46 (0.63)	-0.48 (0.62)	-ve
Edges Rolled Under	-0.86 (0.42)	-	-	-	-0.75 (0.47)	-0.75 (0.47)	-ve
Exudate Amount	0.48 (0.62)	-	-	-	-0.37 (0.69)	-0.39 (0.68)	-ve
Exudate Amount Unknown	1.50 (4.52)	-	-	-	1.11 (3.03)	1.16 (3.20)	+ve
Exudate Type	-0.10 (0.90)	-	-	-	-0.08 (0.92)	-0.08 (0.92)	-ve
Exudate Type Unknown	0.68 (1.98)	-	-	-	0.47 (1.60)	0.47 (1.60)	+ve
Location-1	0.49 (1.64)	-	-	-	0.42 (1.52)	0.41 (1.51)	+ve
Location-3	0.20 (1.22)	-	-	-	0.15 (1.16)	0.17 (1.18)	+ve
Location-4	0.25 (1.28)	-	-	-	0.25 (1.29)	0.26 (1.30)	+ve
Location-5	-0.03 (0.97)	-	-	-	0.02 (1.02)	0.02 (1.02)	Inconclusive
Location-6	0.35 (1.42)	-	-	-	0.37 (1.45)	0.38 (1.46)	+ve
SNF	-0.21 (0.81)	-	-	-	0.09 (1.10)	0.10 (1.11)	Inconclusive

*In the context of the current study, the event of interest is the 'healed' status of the wound (or survival of a wound). Therefore, $b_n > 0$ ($HR > 1$) leads to a positive effect on healing by increasing the wound healing probability, whereas, $b_n < 0$ ($HR < 1$) leads to a negative effect on healing by decreasing the wound healing probability.
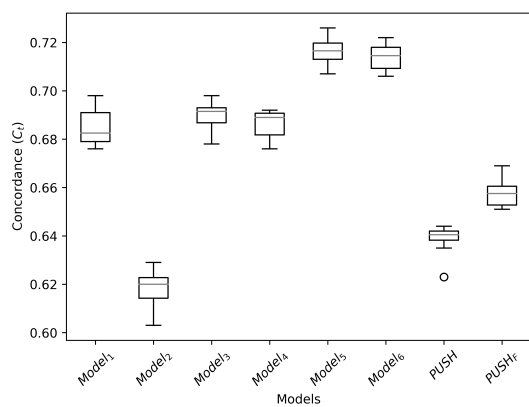

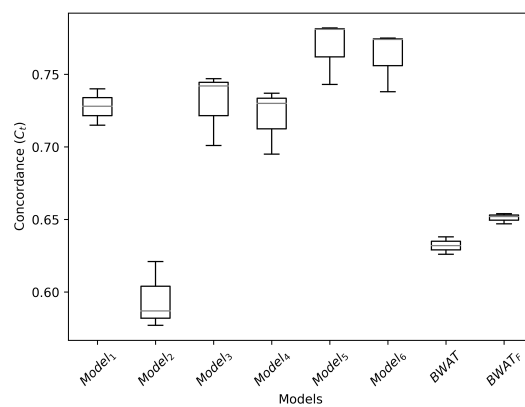Fig. 5: C_t across different models and PUSH score for Pressure Injuries at SNFs.

Fig. 6: C_t across different models and BWAT scores for Pressure Injuries, Venous, Arterial and Diabetic wounds at Homehealth facilities.

1) Dynamic Performance: Figure 7 shows the dynamic $AUC_{I/D}$ for each prognostic model along with standard wound assessment tools used across different clinical settings for pressure injuries. In order to completely characterize the behaviour of the prognostic models across time, the dynamic $AUC_{I/D}$ was computed for an augmented testing dataset. The augmented testing dataset consisted of the original testing data along with wound series that were shorter than three weeks in age. $Model_{5,6}$ consistently performed better than PUSH, forward filled PUSH, BWAT and forward filled BWAT scores across different time points. Generally, each prognostic model performed better than standard wound assessment tools with the exception of $Model_2$, which was trained using subjectively

TABLE VI: Prognostic model comparison across clinical settings using concordance (C_t)

Clinical Setting	Model	All Wound Types	PIs	PIs*	Diabetic	Venous	Arterial
SNF	$Model_1$	0.68	0.68	0.67	0.65	0.66	0.66
SNF	$Model_2$	0.62	0.62	0.60	0.60	0.60	0.63
SNF	$Model_3$	0.68	0.69	0.69	0.65	0.65	0.66
SNF	$Model_4$	0.68	0.69	0.68	0.64	0.64	0.66
SNF	$Model_5$	0.71	0.72	0.71	0.67	0.68	0.68
SNF	$Model_6$	0.70	0.71	0.71	0.66	0.67	0.68
SNF	$PUSH$	NA	0.64	0.62	NA	NA	NA
SNF	$PUSH_F$	NA	0.66	0.64	NA	NA	NA
HHF	$Model_1$	0.73	0.72	0.72	0.73	0.75	0.74
HHF	$Model_2$	0.60	0.57	0.56	0.59	0.63	0.55
HHF	$Model_3$	0.73	0.75	0.75	0.73	0.70	0.61
HHF	$Model_4$	0.72	0.74	0.75	0.72	0.69	0.63
HHF	$Model_5$	0.77	0.78	0.78	0.77	0.76	0.70
HHF	$Model_6$	0.76	0.77	0.78	0.77	0.75	0.70
HHF	$BWAT$	0.63	0.65	0.64	0.63	0.63	0.54
HHF	$BWAT_F$	0.65	0.66	0.66	0.68	0.63	0.64

NA - not applicable, as PUSH is not valid for wounds other than Pressure Injuries; $PUSH_F$ and $BWAT_F$ represent the features computed by forward filling PUSH and BWAT values, respectively, when the information is missing; PIs: Pressure Injuries; PIs*: C_t was computed on the augmented testing dataset consisting longer and shorter than 3 weeks old wounds

identified tissue presence and user traced wound area. The t-tests with Bonferroni-corrected p-values were performed on the SNFs dataset to reveal significant differences between all models across each time point. We found that $Model_{5,6}$ performed better than PUSH and forward filled PUSH scores at least until 150 days since the first evaluation of the wound. Moreover, there was no significant difference between $Model_{5,6}$ across time. Similarly, significant performance differences between $Model_{5,6}$ and BWAT were observed in home healthcare facilities, at least until day 80.

2) Sex-wise: Figure 8 shows the performance of wound-agnostic models across female and male demographic of the population, highlighting that all models performed equally well across sexes. A Mood's median tests performed per model confirmed that there was no significant difference in performance between female and male demographics ($\chi^2 \leq 1.8$, $p > 0.1$).

3) Age-wise: Figure 9 shows the performance of wound-agnostic models across different age groups of the population. Models performed well across different age groups.

D. Wound-Specific Model Performance

The performances of the wound-agnostic models were compared against wound-specific models developed using data from specific wound types. Figure 10 illustrates the performance comparisons between the wound-agnostic and wound-specific models for pressure injuries, diabetic, arterial and venous wounds at SNFs. Mood's median tests performed across model types did not reveal any significant difference between wound-agnostic and wound-specific models for

pressure injuries. However, $Model_{3,4,6}$ showed statistically significant ($\chi^2 = 4.05$, $p < 0.05$) minute differences (up to 1%) between wound-specific and wound-agnostic models for arterial, venous and diabetic wounds.

IV. Discussion

The feature importance analysis identified the most predictive features for wound prognosis, and the most important feature in this analysis was determined to be the amount of exudate. Exudate is a part of the normal wound healing process, however, in chronic and non-healing wounds with persistent inflammation or infection, it leads to delayed healing due to the presence of excess bacteria and abnormal levels of inflammatory mediators and protein digesting enzymes [6]. Furthermore, the remaining features identified as significantly predictive are also directly associated with wound healing. For instance, a firmly attached wound edge has been found to promote wound healing, dark red or beefy-looking granulation tissue within the wound bed is indicative of infection, and wound area is directly correlated with wound closure and healing [7]. Induration and edema were selected during the stepwise selection process, but individually, these features did not account for significant amounts of variations within the data, suggesting that other features encoded a portion of the underlying information represented by induration and edema.

Feature importance information also serves as a guide to build an objective prognostic tool for wound care. Individual machine learning models can be trained to objectively quantify and characterize the relevant features. The resulting objective features can be used downstream to build a prognostic tool through replacing their subjective feature counterparts. To this
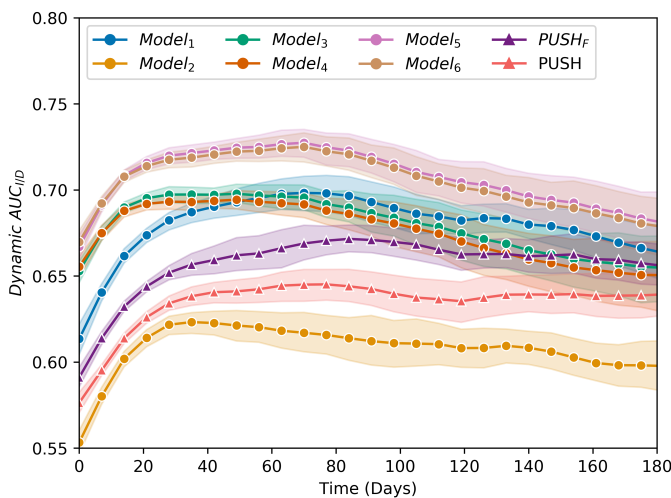
(a) Skilled Nursing Facilities
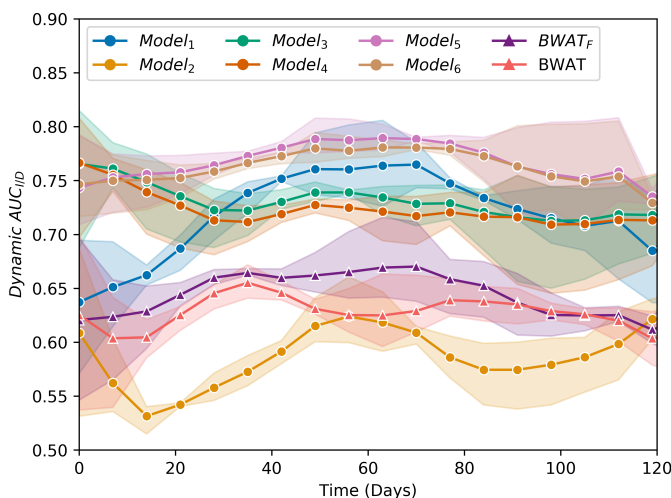
(b) Homehealth Facilities

Fig. 7: Dynamic $AUC_{I/D}$ across different models different clinical settings.
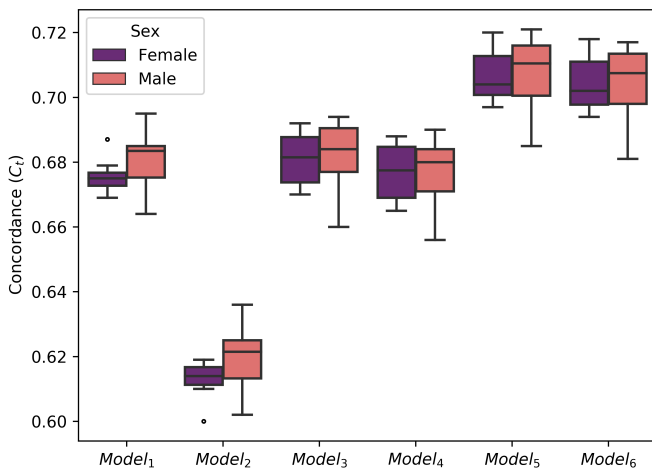

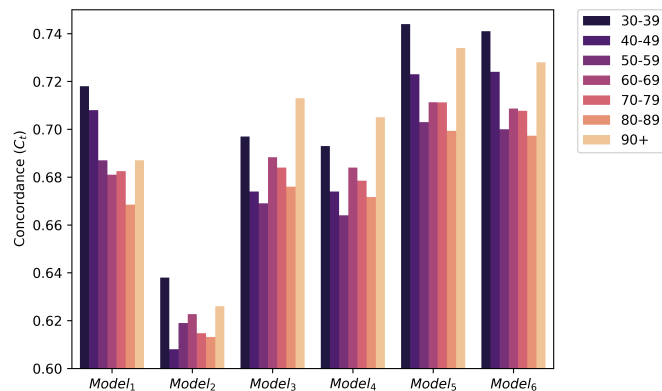Fig. 8: C_t for male and female demographic across different models at SNFs.

Fig. 9: C_t per age group across different models at SNFs. Note that C_t was not computed for age < 30 years due to the absence of sufficient data.

end, the results of this research have shown the feasibility of developing prognostic models using objective representations of *wound extent*. The results from the analysis surrounding feature importance indicate that future research can benefit from focusing on the development of objective representations of factors affecting *wound burden* by modelling and quantifying exudate amount and wound edges.

The developed prognostic models are directly interpretable through hazard ratios corresponding to features constituting a particular model. A feature with $HR > 1$ leads to early wound healing (or decrease in wound survival time), whereas a feature with $HR < 1$ leads to delayed wound healing (or increase in wound survival time). The features with $HR < 1$ that are numerically encoded, such as tissue type presence, normalized wound area, objectively determined granulation, slough, or eschar amounts, exudate amount, and exudate type are features for which an increasing value corresponds to delayed wound healing. Therefore, if necrotic tissue is identified in the wound bed, or wound area increases, or amount of slough tissue increases, or heavy exudate amount is observed at the following wound evaluation, then the risk of delayed wound healing increases. In contrast, features with $HR > 1$ that are numerically encoded, such as other tissue amounts, are features for which an increasing value corresponds to faster wound healing, and thus reduced risk of delayed wound healing. This observation is corroborated by the findings in [21], where the 'other' tissue feature was found to predominantly quantify the amount of healthy tissue. One-hot encoded and boolean features can be interpreted similarly. Therefore, the presence of a feature with $HR < 1$, such as the presence of non-attached or rolled wound edges, or a wound location as on the foot (Location-5), are features that correspond to delayed wound healing. Conversely, the presence of a feature with $HR > 1$, such as the presence of attached or epithelializing wound edges, or a wound location as on the face (Location-1), are features that correspond to faster wound healing.

In this work, a new prognostic model based on subjective features, $Model_1$, was introduced and demonstrated to lead by 4% and 7% over PUSH and BWAT, respectively. Through
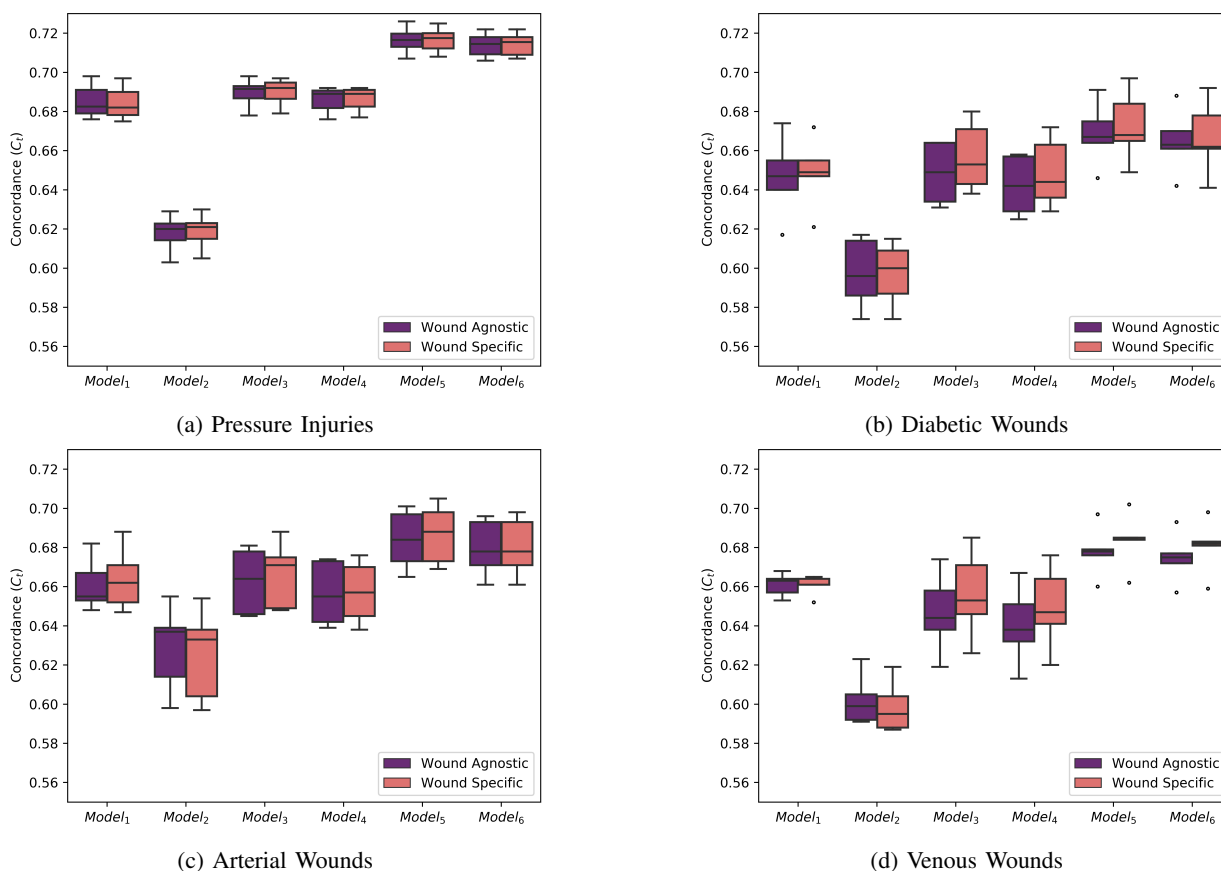
(a) Pressure Injuries

(b) Diabetic Wounds

(c) Arterial Wounds

(d) Venous Wounds

Fig. 10: Comparison of wound-agnostic and wound-specific models across wound types

(a) Risk trajectory of a healing wound
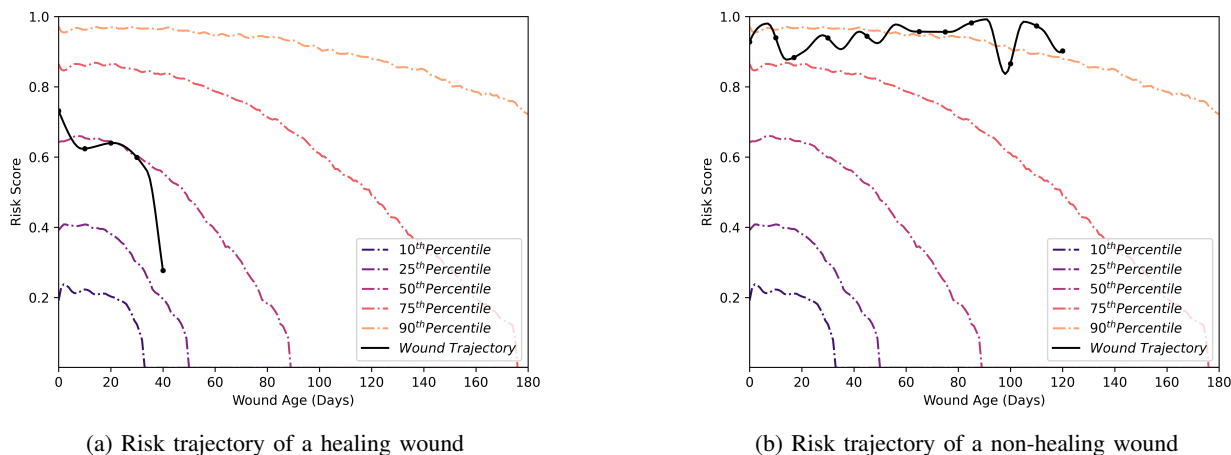
(b) Risk trajectory of a non-healing wound

Fig. 11: An example of the clinical use of the risk or prognostic index derived from $Model_6$. Risk profile percentiles were computed using risk trajectories of wounds within our dataset. Panel (a) represents the trajectory of a wound on track for healing with decreasing risk score during a 40-day period. In contrast, panel (b) shows a non-healing wound with repeated observations consistently above the 90th percentile. For non-healing wounds, our risk model can help in triggering early in-depth wound assessments to identify the reason for non-healing (i.e. infection, use of repair inhibiting drugs, change in dressing type, etc.).

comparing $Model_1$ to $Model_{5,6}$, and $Model_2$ to $Model_{3,4}$, it is evident that replacing subjective information with objective information leads to more accurate prognostic models. The worst performing model, $Model_2$, indicates that using very few subjectively assessed features do not capture the complete information needed to track the progress of wound healing.

The purely objective prognostic model, $Model_4$, consistently performed 6% and 9% better than PUSH and BWAT, respectively, thus providing further validity to the information extracted from the AutoTrace and AutoTissue wound segmentation models. AutoTrace and AutoTissue simply require an image of the wound to compute accurate wound area and tissue

composition, thus $Model_4$ can serve as a faster and more reliable prognostic tool for wound care assessment compared to standard wound assessment tools. Using a hybrid approach such as $Model_{5,6}$ results in an even more accurate prognostic index. A further benefit of the developed models is that they are wound agnostic in nature and have been validated for four different wound etiologies, including pressure injuries, venous, arterial and diabetic wounds. Although, both $Model_5$ and $Model_6$ perform equally well, $Model_6$ would be preferable in practice, as it uses more objective features which are expected to reduce subjectivity and allow quicker assessment of wound severity. The performance of the prognostic models proposed here are expected to be directly correlated with the accuracy of the features derived from the segmentation models. Therefore, it is easy to replace AutoTrace and AutoTissue modules with more accurate segmentation models. However, for most accurate results it would be appropriate to retrain the prognostic models using the new information.

The developed prognostic models performed better than the standard wound assessment tools across two clinical settings, four different wound types, sexes and age groups, establishing their validity across different segments of the dataset. The C_t for the best performing models, $Model_{5,6}$, was at least 0.70 and 0.76, across SNFs and home healthcare settings, respectively. This indicates that the probability of the model-derived risk score being lower for a faster-healing wound than for a longer-healing wound is 70% for SNFs and 76% for home healthcare facilities, assuming that the healed event occurred within the first 180 days of the wound's existence. For home healthcare, this represents a large increase in concordance relative to BWAT, a widely used etiology-agnostic wound assessment tool, which is only about 65% concordant. The differences in concordance across settings also suggest that the models performed better under home healthcare compared to SNFs, though the reason for this observation is unclear. Across SNFs the models performed better for pressure injuries compared to venous and diabetic wounds. This bias could be due to pressure injuries comprising approximately 70% of the total dataset used in training the models. Therefore, wound-specific models were developed by training prognostic models on datasets consisting of specific wound types. Comparing against wound-agnostic models, it was observed that wound-specific models offered statistically significant improvements, though the magnitude of improvement was minuscule (1-1.5%). However, the generalizability of wound-agnostic models across different wound etiologies far outweigh the improvement in performance offered by wound-specific models.

The model performance did not show any significant difference in performance between sexes, thus establishing model generalizability across both males and females. Unfortunately, the model performance could not be characterized across a younger (< 30 years old) population due to insufficient data, as chronic wounds are rare in younger demographics. However, C_t was computed for older populations per age group and the developed models were found to be adequate, with the best performing model resulting in $C_t \geq 0.70$. This establishes the fairness of our prognostic models with regard to demographic

segments of sex and age. The generalizability of our models across different ethnicities remains to be tested in future work.

The model performance was characterized across time using dynamic $AUC_{I/D}$. It was used to compare model performance against standard wound assessment tools used across different clinical settings. The $AUC_{I/D}$ computed for SNFs showed an increase in performance initially up to the first 21-28 days since the first evaluation of the wound, reaching a stable performance state up to 90 days and then slowly decreasing in model performance up to 180 days. Specifically, for $Model_{5,6}$ the $AUC_{I/D}$ ranged between 0.71 and 0.72 within the first 20-120 days of evaluation, indicating that on any day t between 20 and 120 days of evaluation, the probability of a lower risk wound healing faster than a higher risk wound, as predicted by the prognostic models, is at least 0.71. Moreover, we observed that throughout the evaluated time range, our models performed significantly better than the PUSH scores. Similarly, under home healthcare settings, the prognostic models performed better than BWAT scores across time, though the results were not significant after 80 days of evaluation. The lower performance on the initial few weeks of evaluation at SNFs indicates that it is likely difficult to determine wound prognosis for relatively younger wounds using the features used in this study. However, this observation could not be corroborated at home healthcare facilities.

V. FUTURE WORK

The future analysis will explore the feasibility of objectively determining features related to *wound burden*, such as wound edges, exudate amount and exudate types, and their prognostic abilities. In addition, the future models would leverage machine learning techniques, such as random survival forest [44], DeepSurv [45], that model non-linear interactions between features. This will lead to the development of a completely objective, more accurate and faster prognostic tool. We will also explore the generalizability of the reported models across different ethnicities, which is absent from the current analysis due to the unavailability of the information.

Moreover, the current analysis does not consider the impact of prognostic indices on clinical interventions. As future work, we aim to develop a decision support system, as shown in Figure 11, that can assist clinicians in determining the optimal interventions needed to heal the wounds quicker. Given the wound care challenges dependent on skills and experience, being able to suggest clinical interventions that could alter the risk trajectory may provide significant clinical value and make wound care more equitable.

VI. CONCLUSION

In conclusion, this study demonstrates the feasibility of developing prognostic tools based on the objective information derived from wound imaging. The feature importance investigation equips researchers with a guide to develop a more accurate, objective prognostic tool. Furthermore, this study presents superior prognostic models compared to standard wound assessment tools used in the field today across different clinical settings, wound etiologies, sex, age-groups and time.

This indicates that the inclusion of the newly developed prognostic tools into standard wound care practice can aid in accurate and faster detection of high risk wounds, assisting clinicians in better decision making and improving outcomes.

REFERENCES

[1] C. K. Sen *et al.*, "Human skin wounds: a major and snowballing threat to public health and the economy," *Wound repair and regeneration*, vol. 17, no. 6, pp. 763–771, 2009.

[2] G. S. Lazarus, D. M. Cooper, D. R. Knighton, D. J. Margolis, R. E. Percoraro, G. Rodeheaver, and M. C. Robson, "Definitions and guidelines for assessment of wounds and evaluation of healing," *Wound repair and regeneration*, vol. 2, no. 3, pp. 165–170, 1994.

[3] L. McNichol, C. Ratliff, and S. Yates, "Wound, ostomy and continence nurses society core curriculum: Wound management," 2021.

[4] P. E. Houghton, C. B. Kincaid, K. E. Campbell, M. Woodbury, and D. Keast, "Photographic assessment of the appearance of chronic pressure and leg ulcers," *Ostomy Wound Management*, vol. 46, no. 4, pp. 20–35, 2000.

[5] C. T. Hess, "Checklist for factors affecting wound healing," *Advances in skin & wound care*, vol. 24, no. 4, p. 192, 2011.

[6] U. J. Adderley, "Managing wound exudate and promoting healing," *British journal of community nursing*, vol. 15, no. Sup1, pp. S15–S20, 2010.

[7] D. H. Keast, C. K. Bowering, A. W. Evans, G. L. Mackean, C. Burrows, and L. D'Souza, "Contents: Measure: A proposed assessment framework for developing best practice recommendations for wound assessment," *Wound Repair and Regeneration*, vol. 12, pp. s1–s17, 2004.

[8] D. R. Thomas *et al.*, "Pressure ulcer scale for healing: derivation and validation of the push tool. the push task force." *Advances in wound care: the journal for prevention and healing*, vol. 10, no. 5, pp. 96–101, 1997.

[9] N. A. Stotts, G. T. Rodeheaver, D. R. Thomas, R. A. Frantz, A. A. Bartolucci, C. Sussman, B. A. Ferrell, J. Cuddigan, and J. Maklebust, "An instrument to measure healing in pressure ulcers: development and validation of the pressure ulcer scale for healing (push)," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 56, no. 12, pp. M795–M799, 2001.

[10] D. McLaren, P. Med, D. O'Sullivan, and L. Orr, "A prospective, multicenter study to validate use of the pressure ulcer scale for healing (push©) in patients with diabetic, venous, and pressure ulcers," *OSTOMY WOUND MANAGEMENT*, vol. 56, no. 2, pp. 26–36, 2010.

[11] H. Sanada *et al.*, "Reliability and validity of design, a tool that classifies pressure ulcer severity and monitors healing," *Journal of wound care*, vol. 13, no. 1, pp. 13–18, 2004.

[12] D. Krasner, "Wound healing scale, version 1.0: a proposal." *Advances in Wound Care: The Journal for Prevention and Healing*, vol. 10, no. 5, pp. 82–85, 1997.

[13] B. A. Ferrell, B. M. Artinian, and D. Sessing, "The sessing scale for assessment of pressure ulcer healing," *Journal of the American Geriatrics Society*, vol. 43, no. 1, pp. 37–40, 1995.

[14] C. Sussman and G. Suanson, "Utility of the sussman wound healing tool in predicting wound healing," *Adv Wound Care*, vol. 10, no. 5, pp. 74–77, 1997.

[15] F. Chuan, K. Tang, P. Jiang, B. Zhou, and X. He, "Reliability and validity of the perfusion, extent, depth, infection and sensation (pedis) classification system and score in patients with diabetic foot ulcer," *PloS one*, vol. 10, no. 4, p. e0124739, 2015.

[16] P. Ince, Z. G. Abbas, J. K. Lutale, A. Basit, S. M. Ali, F. Chohan, S. Morbach, J. Mollenberg, F. L. Game, and W. J. Jeffcoate, "Use of the sinbad classification system and score in comparing outcome of foot ulcer management on three continents," *Diabetes care*, vol. 31, no. 5, pp. 964–967, 2008.

[17] H. Haeruddin, S. Yusuf, I. Patellongi, K. A. Rasid, H. Harkam, and E. Supardi, "Inter-rater reliability assessment of new diabetic foot ulcers (the diabetic foot ulcer assessment scale) based on photograph," *Enfermería Clínica*, vol. 30, pp. 433–438, 2020.

[18] H. Pillen, M. Miller, J. Thomas, P. Puckridge, S. Sandison, and J. I. Spark, "Assessment of wound healing: validity, reliability and sensitivity of available instruments," *Wound Practice & Research: Journal of the Australian Wound Management Association*, vol. 17, no. 4, pp. 208–217, 2009.

[19] Bates-Jensen *et al.*, "Reliability of the bates-jensen wound assessment tool for pressure injury assessment: The pressure ulcer detection study," *Wound Repair and Regeneration*, vol. 27, no. 4, pp. 386–395, 2019.

[20] S. Stremitzer, T. Wild, and T. Hoelzenbein, "How precise is the evaluation of chronic wounds by health care professionals?" *International Wound Journal*, vol. 4, no. 2, pp. 156–161, 2007.

[21] D. Ramachandram *et al.*, "Fully automated wound tissue segmentation using deep learning on mobile devices: Cohort study," *JMIR mHealth and uHealth*, vol. 10, no. 4, p. e36977, 2022.

[22] W. Paans, R. M. Nieweg, C. P. Van der Schans, and W. Sermeus, "What factors influence the prevalence and accuracy of nursing diagnoses documentation in clinical practice? a systematic literature review," *Journal of Clinical Nursing*, vol. 20, no. 17-18, pp. 2386–2403, 2011.

[23] C. A. Calianno and A. Martin-Boyan, "When is it appropriate to photograph a patient's wound?" *Advances in Skin & Wound Care*, vol. 19, no. 6, pp. 304–307, 2006.

[24] H. T. Mohammed, R. L. Bartlett, D. Babb, R. D. Fraser, and D. Mannion, "A time motion study of manual versus artificial intelligence methods for wound assessment," *PloS one*, vol. 17, no. 7, p. e0271742, 2022.

[25] S. C. Wang, J. A. Anderson, R. Evans, K. Woo, B. Beland, D. Sasseville, and L. Moreau, "Point-of-care wound visioning technology: Reproducibility and accuracy of a wound measurement app," *PloS one*, vol. 12, no. 8, p. e0183139, 2017.

[26] R. D. Fraser *et al.*, "Analysis of real-world data from north american skilled nursing facilities' skin and wound records for skin tear prevalence, healing and treatment," *Journal of Wound Management*, vol. 13, 2022.

[27] C. Wang *et al.*, "Fully automatic wound segmentation with deep convolutional neural networks," *Scientific reports*, vol. 10, no. 1, pp. 1–9, 2020.

[28] G. Scebba *et al.*, "Detect-and-segment: a deep learning approach to automate wound image segmentation," *Informatics in Medicine Unlocked*, vol. 29, p. 100884, 2022.

[29] H. Carrión *et al.*, "Automatic wound detection and size estimation using deep learning algorithms," *PLoS computational biology*, vol. 18, no. 3, p. e1009852, 2022.

[30] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," pp. 3431–3440, 2015.

[31] R. B. Kim *et al.*, "Utilization of smartphone and tablet camera photographs to predict healing of diabetes-related foot ulcers," *Computers in biology and medicine*, vol. 126, p. 104042, 2020.

[32] C. Wang *et al.*, "A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks," pp. 2415–2418, 2015.

[33] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972.

[34] D. Curran-Everett, "Explorations in statistics: the log transformation," *Advances in physiology education*, vol. 42, no. 2, pp. 343–347, 2018.

[35] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," pp. 6105–6114, 2019.

[36] S. L. Spruance, J. E. Reid, M. Grace, and M. Samore, "Hazard ratio in clinical trials," *Antimicrobial agents and chemotherapy*, vol. 48, no. 8, pp. 2787–2792, 2004.

[37] P. Royston and D. G. Altman, "External validation of a cox prognostic model: principles and methods," *BMC medical research methodology*, vol. 13, no. 1, pp. 1–15, 2013.

[38] L. D. Fisher, D. Y. Lin *et al.*, "Time-dependent covariates in the cox proportional-hazards regression model," *Annual review of public health*, vol. 20, no. 1, pp. 145–157, 1999.

[39] C. Davidson-Pilon, "lifelines: survival analysis in python," *Journal of Open Source Software*, vol. 4, no. 40, p. 1317, 2019. [Online]. Available: https://doi.org/10.21105/joss.01317

[40] F. E. Harrell *et al.*, *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer, 2001, vol. 608.

[41] F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati, "Evaluating the yield of medical tests," *Jama*, vol. 247, no. 18, pp. 2543–2546, 1982.

[42] N. van Geloven, Y. He, A. H. Zwinderman, and H. Putter, "Estimation of incident dynamic auc in practice," *Computational Statistics & Data Analysis*, vol. 154, p. 107095, 2021.

[43] P. J. Heagerty and Y. Zheng, "Survival model predictive accuracy and roc curves," *Biometrics*, vol. 61, no. 1, pp. 92–105, 2005.

[44] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer, "Random survival forests," *The annals of applied statistics*, vol. 2, no. 3, pp. 841–860, 2008.

[45] J. L. Katzman *et al.*, "Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network," *BMC medical research methodology*, vol. 18, no. 1, pp. 1–12, 2018.