A Comprehensive Methodology for Enhanced Wound Care: Integrating Machine Learning to Predict Patient Demographics, Wound Healing Status and Trajectory

Atika Rahman Paddo[1], Arushi Tibrewal[1], Shikhar Shukla[1], Shomita S. Steiner, Saptarshi Purkhayastha[1], and Chandan K. Sen

School of Informatics, Computing and Engineering
Indiana University Purdue University Indianapolis, IN, USA
apaddo@iu.edu

Abstract. Chronic and non-healing wounds present a global challenge that requires specialized treatment in wound clinics. The high amputation rates in Indiana emphasize the urgent need for effective wound care programs that improve efficiency, showcase innovation, and enhance patient outcomes. To address these needs, DHIS2-WoundInfo, an open-source web-based platform is being developed for comprehensive wound data collection, management, and analysis. It utilizes predictive analytics and advanced machine learning models to advance value-based wound care and research. The platform stores, integrates, and analyses wound data, providing valuable insights into healing patterns, demographic factors, and evidence-based delivery planning for physicians and researchers. The primary objective is to conduct an in-depth analysis of wound data obtained from the merged and integrated Wound-Expert database. Through machine learning models, wounds are classified into five different classes. The aim of this is to help doctors devise improved/specialized treatment plans for patients with non-healing wounds, to reduce amputations due to ineffective and untimely wound treatment.

Keywords: ... · ... · ... · ....

Abstract. Chronic wound treatment and management heavily rely on accurate patient outcome predictions. As chronic wounds become increasingly prevalent in many US states, it's crucial to accurately predict and classify wound progress for timely and effective clinical intervention. The trajectory and outcomes of wounds inform multiple aspects of care, from estimating healing times to understanding disparities in treatment quality and outcomes. For accurate predictions, robust time-series forecast models are required, necessitating comprehensive and consistent wound data. Yet, much of this data is frequently marred by gaps and inaccuracies. In this work, we combined two approaches to overcome this challenge: filling in missing clinical data and forecasting wound healing trajectories. We employed various interpolation techniques, time series, and machine-learning models, including Holt-Winters, ARIMA, Prophet, LSTM, and BiLSTM. Our study assessed 14,571 wounds from 6,171 patients over a five-year span, post-data imputation and interpolation. We introduced a novel approach to firstly impute missing data using XGBoost and then choose the best interpolation method for absent wound data, ensuring precise forecasting without distorting the dataset's demographic makeup. Data for this research was sourced from four wound centers in XX and eight research studies, focusing on a diverse range of wound-related demographics, clinical, and omics data. This comprehensive approach enhances wound trajectory predictions, providing clinicians with valuable insights to formulate personalized treatment plans for both individual patients and demographic groups. This will aid clinicians in better understanding wound healing trends and crafting targeted treatment strategies.

Keywords: Predictive models, Health informatics, Data analysis, Information analysis, Forecasting, Wounds, Health care

1 Introduction

Chronic wounds, failing to heal in a predictable amount of time, represent a significant and growing health concern, especially in developed countries. The prevalence of chronic wounds, including diabetic foot ulcers, venous leg ulcers, and pressure ulcers, is on the rise. A report by [30] highlighted the increasing threat of chronic wounds, particularly in the United States, where millions are affected with associated medical costs in billions. Accurate determination of a wound's healing trajectory can markedly influence clinical decisions, potentially optimizing treatment outcomes. [8] highlighted the importance of understanding the underlying pathophysiology and managing factors that contribute to delayed healing to achieve successful healing. With the advancement of computational methods, Machine Learning (ML) models present an avenue with significant promise in refining our predictions regarding wound healing trajectories. For example, [26] used ML methods and attribute estimation algorithms to rank prognostic factors and build models for predicting the wound healing rate.

However, the challenge before making a good prediction is that of data quality. As observed by [12], missing data can be multifactorial, leading to inaccurate and inconsistent predictions. Thus, it's essential to develop techniques to rectify and interpolate these data, ensuring the most reliable foundation for predictive models. In this paper, we combine ML approaches for both predicting wound healing trajectories and refining underlying data through interpolation techniques, aiming to provide clinicians with robust insights for improved wound care.

The investigation of wound progression and outcomes has gained considerable prominence in recent years. Wound care is an interdisciplinary domain that strives to deepen the understanding of wound healing processes, pinpoint elements affecting wound care results, and devise innovative strategies to refine the healing process ([31]). Many complex processes involved in wound healing rely on the interaction of multiple time-dependent components ([27]), which may or may not require multivariate forecasting ([23]). Precise estimation of wound healing duration will enable clinicians to customize treatment approaches and allocate resources effectively, thus augmenting patient outcomes ([10]). Factors such as socioeconomic status, geographic location, and health insurance provisions and addressing disparities in care access, can enhance wound care quality and patient satisfaction ([14]). Furthermore, [17] showed that disparities in wound care services, such as differences in access, quality, and outcomes, can contribute to health inequalities among diverse population groups. By identifying and addressing these disparities, researchers and practitioners can develop targeted interventions to ameliorate wound care outcomes for susceptible populations ([18]). Recent work on COVID-19 resource usage shows that demographic details are essential to make disease trajectory predictions ([36]). Thus, having complete demographic data is critical for disease trajectory prediction. This article describes a novel ML-based approach to impute missing demographic data and interpolate lost-to-follow-up wound measurement. We provide evaluation results to demonstrate the usefulness of our method that can be used to improve wound trajectory and forecasting.

We developed our work to use the data imported to DHIS2, a popular open-source data warehousing platform, with patient-specific reports that use models to forecast when and how a wound will heal based on wound assessment or visit information. To validate our work, we used data from Net Health Wound Care (formerly WoundExpert) ([20]) to train the models with a large number of data values.

(Extra) According to the U.S. Census Bureau, approximately 56 million people, which is over 18 of the noninstitutionalized population, had a disability in 2010 (Brault, 2012). The main objective of this project is to reduce what is referred to as "late amputations", however, there is no consensus definition for "delayed" or "late" amputation. Authors' definitions of late amputation range from 24 hours to more than a year following the injury (Stinner et al., 2010).

Amputation poses a major health burden on families, society, and medical services. Traumatic limb amputation is a catastrophic injury, an irreversible act that is sudden and emotionally devastating for the victims. Furthermore, it leads to an inability to support oneself and the family, often resulting in various psychiatric disorders (Sahu et al., 2016).

Non-healing wounds affect about 3 to 6 million people in the United States, with persons aged 65 and older accounting for 85 of these cases. Non-healing wounds also result in enormous healthcare expenditures, with the total cost estimated at more than 3 billion per year (Menke et al., 2007). Therefore, measures need to be taken to reduce disability. Efforts should be made to reduce delay before hospital presentation, improve first-aid knowledge, enhance wound and clinical care in communities, and upgrade facilities (Abubakar et al., 2010).

The endeavor of DHIS2-WoundInfo is to provide the necessary information to help achieve these goals. By targeting delayed wound treatment and employing data-driven approaches, the aim is to significantly reduce unnecessary limb amputations.

Wounds traditionally are classified manually by wound specialists and made as part of the electronic health record (EHR). With new developments of Artificial Intelligence (AI) in the past decades, however, intelligent algorithms have been popularly used in healthcare in such fields as drug discovery, eye care, medical image diagnostic systems, etc (Yu et al., 2018). In a recent study, researchers trained a DCNN (Deep Convolutional Neural Network) to classify chronic wound images into different types of ulcers based on a dataset of wound images. They achieved a 92 accuracy using this DCNN model for their 4-class classification results (Rostami, 2021). In another study, the use of random forest and SVM (support vector machines) was analyzed for wound image evaluation where the researchers achieved high performance rates. According to the study, support vector machines and random forests gave the high performance rates when classifying wound-bed patterns composed of color, texture, region morphology and topology features extracted from the segmented regions in a set of real pressure ulcer images (Veredas, 2015). Healthcare applications based on AI are utilized in early detection, diagnosis, treatment, as well as outcome prediction and prognosis evaluation (Jiang et al., 2017).

2 Related Works

As our initial work revolved around using ML methods to overcome the problem of finding chronic wound trajectory over the period of time a patient visits a care center, we explored the use of ML-based algorithms in overall wound care research. [4] examined the prediction of patient outcomes following severe injury, revealing accurate predictions using the SuperLearner algorithm. [25] utilized the WoundFlow software to map burn wounds and identify factors affecting healing, finding associations with age, burn severity, and renal dysfunction. [15] focused on predicting open wound size, achieving high accuracy with a combined ML model. [21] investigated machine learning classifiers for wound care decisions, highlighting improved accuracy with clinical expert input. [5] explored wound healing rate prediction using regression trees and a mathematical model. These studies demonstrate the diverse applications of wound trajectory research, including forecasting healing time, improving access to care, addressing disparities, and enhancing outcomes. Effective record management systems play a crucial role in supporting such research endeavors.

Assessment of wound healing may be subjective, where a healed/healing wound can be differentially defined based on the purpose of the research or clinical study ([40] & [7]). Understanding the wound healing trajectory from each patient level data can be used in precision medicine with a consensus that each patient's wound healing trajectory tends to differ. Therefore, we have objectively tried to show each patient's wound healing trajectory and predict their future healing process from the forecasting models.

Table 1: Baseline Wound Trajectory Models' Performance

	Baseline for REDCap Research Wounds		**[28]**	**[24]**
Model	LSTM	Auto-ARIMA	Linear Regression	Linear Regression
# wounds	320	380	83	171
Average R^2	1	0.94	0.98	0.97

It has been seen from several studies that predictive models built on ML approach can be effective in predicting medical and omics data such as works by [9], [2], [41] and [19]. The wound healing trajectory result was stratified based on the patients' features, such as gender, race, and ethnicity. As we had many missing information in these features, we applied Missing Value Estimation methods to impute the missing values. [39] and [38] have discussed about kNNImpute method and different algorithms to improve the efficiency of the imputation, respectively. From a correctly interpolated dataset, getting to know future stratified wound healing trajectory from an ML model, healthcare providers can largely benefit from seeing the trajectory of a wound; alternative patient care or treatment methods can be developed. Thus machine learning approaches could enable visualization of a patient's wound trajectory to help identify wound care measures that could better benefit the patient in the long run.

Time-series forecasting models, while a powerful tool for predicting future events based on past data, are not immune to the biases that can be inherent in their training data, especially when this data reflects systemic inequalities related to race, ethnicity, or gender. For instance, predictive policing algorithms, which often utilize time-series data to forecast crime rates and deploy resources accordingly, can perpetuate racial biases present in historical arrest data. This leads to a vicious cycle where minority neighborhoods are over-policed, skewing future data and predictions [16]. Another example in healthcare is using algorithms to predict future patient needs. [22] found that an algorithm widely used to guide care decisions was less likely to recommend additional medical assistance for Black patients than for White patients with similar health conditions. This was because the algorithm was trained on cost data, and not actual patient needs, reflecting existing disparities in healthcare spending. Gender bias can also creep in; for instance, when forecasting purchasing behaviors, past data might reflect gendered buying patterns, leading models to pigeonhole consumers into stereotypical categories [6]. It's crucial to be aware of these pitfalls to ensure that time-series forecasting models are both accurate and equitable. Thus, for every model that we built in this project, we performed subgroup analysis by gender and race-ethnicity (see Table 4) before deployment.

Our initial wound trajectory forecasting was done on eight research studies from REDCap, which we consider baseline models. Literature shows that linear regression achieved good results in clean, well-organized data in clinical trials in controlled groups. However, linear regression models (GEE/GLM) did not often perform well on incomplete data. Forecasting performance from baseline and related work is shown in Table 1.

3 Methodology

3.1 Data Collection

To develop ML models and time-series forecasting of wound trajectory, we retrieved data from WoundExpert hosted at XX. The wound data contained 1173 data tables. To navigate through the data and accelerate the data analysis process, we created a data dictionary for the elements in the data tables.

3.2 Building a Framework of Large Wound Registry Data

The study received approval from the Institutional Review Board (IRB) at XX. When analyzing data tables from WoundExpert, it was evident that many did not provide the necessary information for our wound forecasting models. Out of 1,161 tables, a significant number were either null, incomplete, or lacked essential data columns, such as wound/document ID or patient ID.

Furthermore, wound assessments from various clinics demonstrated inconsistencies in the data quality needed to develop a robust forecasting model. Some examples included inconsistent wound assessment data for consecutive dates or varying assessment details on the same date. The data also contained outliers which were subsequently removed. To enhance data quality and address these inconsistencies and outliers, we employed multiple interpolation and forecasting models for all patients in WoundExpert.

3.3 Wound Healing Status Classification

- We started with making a large merged dataframe from the separate tables from WoundExpert database.
- How the merging thing worked:
  - We matched data elements which are mostly common in both REDCap and WoundExpert dataset and selected the tables which contained the matched data elements from WoundExpert.
  - The tables from WoundExpert that were taken for merging, Protective Health Information from those tables were removed the first time so that the tables can be used independently for further analysis.
  - The relevant data elements from those different tables (without the PHI data) are marged. Each data elements were merged together using Python's dictionary approach.
- After we have a big merged dataframe containing 210 data elements, we began data preprocessing. How the data preprocessing worked:

- We started by dropping columns that were not relevant for the classification model. Information like indexes, document types, different kinds of IDs, and image file count, were some columns that were dropped. Columns that had no relevant information or ambiguous information were also dropped. All the irrelevant data elements which will not be used in training the classification model have been filtered out from the dataset. The irrelevant data elements include mostly the IDs which are not PHI, varies facility to facility and are not necessary for training a model. These data elements are Woundassessment_UserID, patientmedications_DocumentType, patientmedications_VisitID, patientmedications_UserID, Wound_DocumentType, Wound_UserID, Wound_PhysicianID, AdmissionID, previousID, SyncUserID, ReferringPhysicianID, patientmedications_Title, patientmedications_PhysicianID, Woundassessment_PhysicianID, Patient_UserID, PhysicianID, Woundassessment_VisitID. That returned 190 columns.
- Then we check how many columns or features or data elements from the dataset have how many non-null values. And if any relevant features from the dataset have less than 5000 non-null data values, we dropped/removed the feature from the dataset because having very less number of data values will not contribute much in further analysis. That returned 182 columns.
- In this filtered dataframe, we had wound information of total 8728 patients having a total of 25255 unique wounds.
- For all the date-related columns (ie. Wound Assessment/Visit date, When the wound information was first recorded, when the patient was first admitted into the facility, Date of Birth and Date of Death and all), we did a modification. We checked the first assessment date for each unique wound. Then we calculated the Day number for each date from the first assessment date for the patient's specific wound. For example, if a patient visited the facility for the 3rd time to have their wound assessed and the first assessment date was on July 1st 2021, 3rd visit date was on July 17 2021, we considered first assessment as Day 1 and Visit 3 as Day 17. Further if the patient visited the facility for a next visit on August 1st, we considered Visit 4 as Day 32. For deceased patients, their date of death was also modified like this as Day number from First Assessment Date.
- For each of the assessment or Visit, we created the ground truth of the wound healing status. The ground truth i.e. the wound healing status was created as following:
  * For each visit, the wound's area value is observed. It is compared to the previous visit's area value. Then the wound's closure from the previous visit is compared.
  * If the wound closure from the previous visit is 100%, then it is considered as "Healed" wounds. If it more than 65% closed, it is considered as "Healing" wound, if the closure is between 20-65%, it is in "Grey Zone", we neither can say if it is healing or not healing. If the closure is less than 20%, then it is considered as "Not healing"

or "Non-healing". If there is no wound closure or the wound area has been increased, then for sure the wound is not healing. This is how the ground truth for the classification model to predict wound healing status is created.

  * As we created the wound healing status as ground truth, new data elements in the dataset are added such as Day Number, Wound Healing Percentage and Wound Healing Status

- After creating the ground truth, the data elements are checked and categorical and numerical variables from the dataset are defined. The categorical variables are label encoded later.
- LightGBM and XGBoost models are used to get feature importance from the data elements to predict Wound Healing Status. If anty feature which have 0 feature importance value according to these models, those are removed and finally we have 134 importance features in the dataset.
- Correlation is checked for each of the remaining features of the dataset and if any two features are highly correlated (either negatively or positively correlated), we check their feature importance according to LightGBM model on predicting the Wound Healing Status, and remove the less importance feature among the two highly correlation features. As we have different features in this dataset, and as we have both categorical and numerical variables in it, we used different method of finding correlation coefficient depending the data distribution of the numerical variables and used point biserial correlation method for categorical variables.
- After removing the features from feature importance and correlation, we had 117 data elements in the filtered dataset.
- Then we transposed the dataframe to make this dataset suitable for classifying each unique wound's final wound healing status. The transposing tasks were as follows:
  * Based on the unique WoundID number for each patients and each wounds, we obsered the highest number of visits the patient had for specific wounds and the variable information for each visits are considered as separate features and the row-wise data for each visit are transposed and added in separate columns. It means, for one unique wound, each visit's wound area value, length, depth, wound color, healing status are considered separate features or columns in the dataset. The non-varying data such as the height, race or gender information however are kept unchanged and kept as single features in the dataset. For each wounds, we did this transpose and most of the patients have multiple visits, the number features in the dataset also got increased in terms of visit numbers. The highest number visit one patient had 469. The varying features or information for each visit were 66, non-varying features were 49. Thus in the final transposed dataframe, we had 49+469*66=31003 columns for 24437 unique wouns. Initially we had 25255 wounds which were duplicates

of some wounds. After removing the duplicated, it came down to 24437 unique wounds.
    * The transposed dataframe also contains a new column which is "FinalVisitHealingStatus". As our goal is to predict the wound healing status on the final or last visit day, we took out the last visit day's healing status as ground truth for our further classification model and make a new column in the dataset. Thus we had finally 31004 features for each wounds.
  - The dataset contained having different which have total number of visits ranging from 1 to 469. As we want to predict the healing status which is derived from at least on previous visits, the wounds having only one visit will not make sense in it. Thus we removed those wounds which do not have information after the first visit, meaning those patients did not have any assessment information for visit 2 and so on for their specific wounds. It was observed there are 4509 number of unique wound of such kind, thus we had 24437-4509=19928 wounds for which have 2 or more visit/assessment information and trained the final classification models.
- The classification model training tasks are as follows:
  - As we have a lot features to be added in the training process, we needed to do feature reduction. Most of the unique wounds do not contain a lot visit information, thus we dropped/removed those visit information/features where we do not have information for 90% or more wounds. Thus the feature got reduced from 31004 down to 1200. All other features which got removed had information for less than or equal to for 10% total wounds, thus they got removed.
  - All of these features (=1200) are observed and categorical and numerical variables are defined and categorical variables are label encoded.
  - The way we reduced the features before making the transposed dataframe, same techniques were applied to again reduced the features: feature importance and correlation checking (Add detail Atika)
  - Once we removed the less important features according to feature importance and correlation check, the dataset came down to 467 features along with the target variable of final wound healing status. These features are mostly comprised of varying features from visit 1 to visit 18.
  - The values of each column were manually checked to infer the type of column: categorical or continuous(numeric). Column types were manually set to "category" as applicable. Numeric columns did not need to be changed.
- 467 features after feature reduction are used in training the LightGBM and XGBoost model for classifying the wound healing status for final visit.
- We want to use the important features in the forecasting as well. The important features according to feature importance from LightGBM model that can be used in the Multivariate Forecasting model are sorted based on their importance: Wound Healing Percentage, Wound Width, Wound Length, Wound Healing Status (During every visits), Wound Depth. These should be incorporated in the forecasting model one by one.

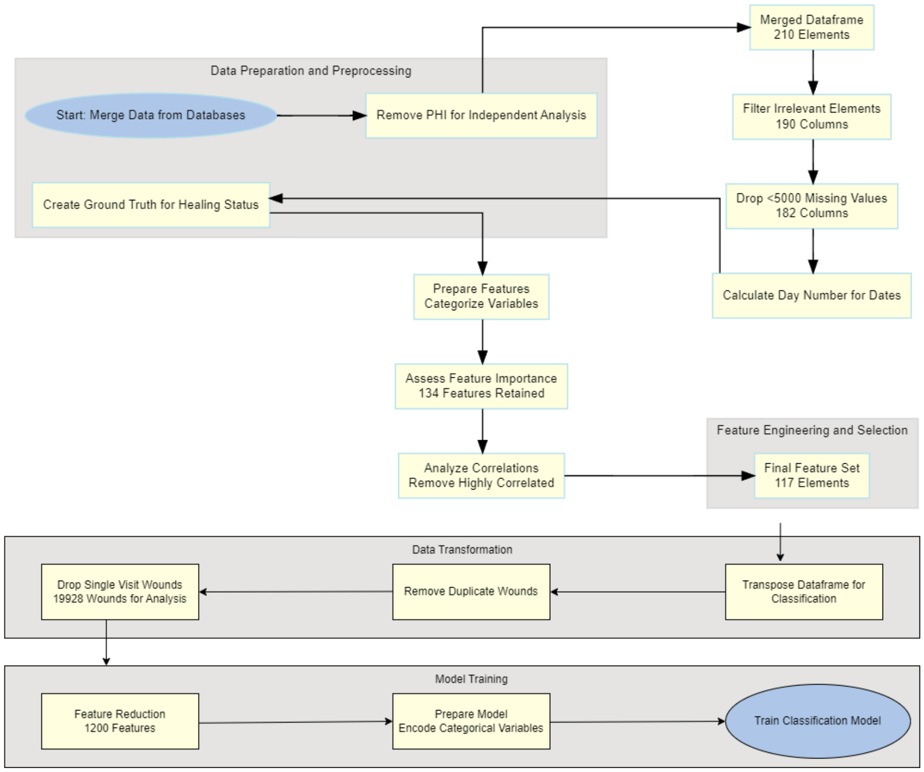


Fig. 1: Caption for the image.

3.4 Forecasting of Wound Healing Trajectory

We observed the important features from the classification model which we want to use in the time series forecasting model. We extracted those features from different data tables from WoundExpert and merged those together. Along with those, as we want to categorize the result based on the gender, race and ethinicity information of the patients and see if there is any bias in wound healing among patient demographics, we included the demographic features in the merged dataset.

Data Imputation To stratify our forecasted model performance by categories like gender, race, and ethnicity, we needed to fill in missing information for many patients. We chose the XGBoost Classification, an ML-based method, to estimate these missing values. For training this model, we used data like age, smoking status, BMI, wound measurements, number of wounds, and the existing gender, race, and ethnicity details from a combined data frame. For fine-tuning the model, we made the following adjustments: Maximum depth of each tree: 12; Number of features considered at each split: log base 2 of the total features; Maximum number of leaf nodes: 3 and Total number of decision trees: 25. These settings determine the model's complexity, the way it selects features, and the total trees in the ensemble. Once trained, the model helped us fill in the missing gender, race, and ethnicity details for some patients.

Table 2: Gender, Race and Ethnicity Distribution before and after Imputation using XGBoost Classifier

	Before Imputation			**After Imputation**		
Gender	*Count*	*Percentage*	*Unknown Gender Count*	*Count*	*Percentage*	*Unknown Gender Count*
Male	3833	51.87%	10	3841	51.91%	0
Female	3556	48.13%		3558	48.09%	
Race	*Count*	*Percentage*	*Unknown Race Count*	*Count*	*Percentage*	*Unknown Race Count*
White	3240	83.40%	3514	6277	84.84%	0
Black	588	15.14%		885	11.96%	
Asian	24	0.62%		52	0.70%	
ASKU(Ask but Unknown)	22	0.57%		147	1.99%	
American Indian/ Alaska Native	6	0.15%		9	0.12%	
Native Hawaiian/ Other Pacific Islander	3	0.08%		20	0.27%	
Other	2	0.05%		9	0.12%	
Ethnicity	*Count*	*Percentage*	*Unknown Ethnicity Count*	*Count*	*Percentage*	*Unknown Ethnicity Count*
Non-Hispanic	4930	97.99%	2368	7222	97.61%	0
Hispanic	92	1.83%		168	1.83%	
Patient Declined	8	0.16%		8	0.11%	
ASKU	1	0.02%		1	0.01%	

After this process, we compared data distribution before and after imputation. There was a minor difference, so we conducted the Kruskal-Wallis test. The results, with a test statistic of 0.92 and a p-value of 0.34, showed no significant difference in data distribution before and after the imputation. Table 2 shows distribution before and after imputation.

Data Cleaning Few patients had only one wound assessment. We excluded these patients because a single assessment doesn't allow for any predictions about the wound's past or future trajectory. Next, we discarded data for patients with multiple wounds when specific wounds only had one assessment. Ultimately, our refined dataset only included wounds with at least two assessments in the WoundExpert database.

During the final step of data interpolation, some wounds, despite having multiple assessments, only had between 2 to 5 total visits. This small number of assessments made validation on a test set challenging. Thus, we categorized these as "Lost to Follow-up" wounds due to their limited visits and removed them from further analysis.

Interpolation We used various interpolation methods to fill in the missing data for wound measurements. However, since we've identified three primary patterns in wound area distribution - Radial, Polynomial, and Linear - we have chosen to focus on these three interpolation types. For polynomial types, we have chosen Akima1D and Krogh interpolation methods.

3.5 Methodological Improvement in the Interpolation

Before diving into forecasting, we applied various interpolation methods to our preprocessed data. For our forecasting models, which can be time-series or neural network-based, it's essential to have a comprehensive dataset. This meant that any missing assessment data in the wounds had to be interpolated. We found that relying on just one interpolation technique wasn't effective for many of the wound measurements. Recognizing this, we suggested a more flexible approach to selecting interpolation methods. We noticed that inaccurate interpolation led to poor outcomes when using the time-series models. Some interpolation methods have constraints, like the Akima1D method, which requires at least five data points.

To address this, we tried interpolating using three different types of wound assessment data: Radial, Polynomial, and Linear. We then evaluated the results using the Akaike Information Criterion (AIC) to determine the most suitable method for each wound. Among the Polynomial interpolation methods, we specifically looked at Krogh and Akima1D, given their popularity. For each wound, the interpolation method that yielded the lowest AIC was chosen to complete the missing data. For those interested, the code detailing our data cleaning, merging, interpolation, and forecasting model is available here: `https://anonymous.4open.science/r/wound-forecast-0271/`.

(to be added more by Atika for forecasting)

Table 3: Interpolation Result on The Wounds

		Good Interpolation RMSE<5 MAPE<5		**Bad Interpolation RMSE>5 MAPE>5**		**Total Interpolation**	
Number of wounds		*n=14417*		*n=154*		*n=14571*	
		Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Interpolation Method Validation Metrics	***AIC***	294.07	946.65	3576.05	3722.80	328.76	1069.98
	RMSE	0.45	0.92	44.88	60.49	0.92	7.74
	MAPE	0.20	0.51	300.80	1867.02	3.38	193.80
	R^2	0.84	0.27	0.46	0.32	0.83	0.27
Interpolation Method Distribution	***Linear***	65.35%		82.47%		65.53%	
	Krogh	15.47%		1.30%		15.32%	
	Akima1D	13.41%		5.84%		13.33%	
	RBF	5.77%		10.39%		5.82%	

Time-Series Forecasting Models We used 5 forecasting models to forecast the future wound healing trajectory - Holt-Winters, ARIMA, Prophet, LSTM, and BiLSTM. Each of the models was suitable for certain types of wounds. Each model was run on the interpolated data of the individual wounds.

Holt-Winters Holt-Winters is a well-known model to see the time series behavior of data. Through Holt-Winters, we can see three aspects of time-series data which are average, slope or trend, and cyclical repeating pattern or seasonality.

Auto-ARIMA An Auto-Regressive Integrated Moving Average (ARIMA) model is an advanced way of predicting time-series data ([33]).

Prophet Prophet was first introduced by Facebook and proposed by [37]. As Prophet was originally proposed for forecasting daily data with weekly and yearly seasonality and was later extended for seasonal data as well, we opted for it to get the forecasted wound trajectory as well.

LSTM LSTM, or Long Short-Term Memory, is a special type of Recurrent Neural Network (RNN) and can be enabled to learn long-term dependencies in data. LSTM is used in time-series forecasting because of the spatio-temporal nature of wound data. An LSTM network learns from a sequence of past data as input and presents an output observation. Several researchers have used LSTM networks for forecasting time-series data [42], [1], [29] and [13].

Bi-LSTM Bi-LSTM or Bi-Directional Long Short-Term Memory uses two independent LSTM networks together. Thus the network has both backward and forward information from the sequence while training the model. Bi-LSTM has shown promising results in time-series forecasting, particularly in data that has cyclical or multi-directional characteristics [34], [35] and [32].

4 RESULTS

4.1 Merged Data Frame & Imputation by XGBoost Classifier

The final merged data frame from six data tables gave us a dataset having 1,726,834 rows and 11 columns. The 1,72,6834 rows represented the number of total wound assessments, and 11 columns represented the features, including Patient Identification Number, corresponding Date of the assessment (Converted from Date-Time format to Date format for simplicity), Wound Number, Length and Width information of the wound on the assessment day, Patient's gender, Race, Ethnicity, Age, BMI, and Smoking Status.

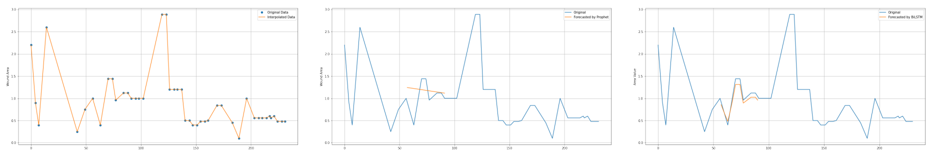
Before imputing gender, race, and ethnicity, there were 3,833 known male patients, 3,556 known female patients, and missing values for 10 patients' gender, 3,514 patients' race, and 2,368 patients' ethnicity. We filled in these missing values using a machine learning-based imputation method (3.4). The results are shown in Table 2.

4.2 Interpolation

After data cleaning, we found 14,918 wounds for which interpolation could be done (having more than one assessment). But some wounds (n=347) were considered Lost to Follow-up wounds as the total number of visit days after interpolating these wounds was much less (2/3/4/5 days); thus, we couldn't validate such data. So we finally analyzed 14,571 wounds. Each interpolation method (RBF, Krogh, Akima1D, Linear) was applied to each wound's data, and the AIC values for each fitted interpolation function were noted. For each wound, we selected the interpolation method with the lowest AIC value as the final interpolation and noted the interpolated values. To verify the technique, we re-interpolated the data for each wound, repeating the method of switching to the best-fitted interpolator again. We kept random 20% sequential data of the interpolated data from the first interpolation for testing and used the remaining data (80% of the total data) to do the re-interpolation. We evaluated the results by calculating RMSE, MAPE, and R^2 values for the removed data (Test set) and interpolated data (Predicted set) from the second interpolation. The average values of the metrics are shown in Table 3.

Table 4: STRATIFIED AVERAGE ACCURACY METRICS FOR THE FORECASTING MODELS

Models			**Holt-Winters**	**AutoARIMA**	**Prophet**	**LSTM**	**BiLSTM**
Features for Stratified Result		**Accuracy Metrics**	***Average (±SD)***	***Average (±SD)***	***Average (±SD)***	***Average (±SD)***	***Average (±SD)***
Gender	*Female*	*RMSE*	14.71 (±75.04)	4.84 (±48.64)	**2.58 (±15.20)**	4.84 (±24.85)	6.04 (±28.30)
		R^2	0.78 (±0.30)	0.86 (±0.25)	0.84 (±0.27)	0.86 (±0.18)	**0.93 (±0.17)**
	Male	*RMSE*	20.14 (±151.83)	6.17 (±88.88)	**3.07 (±18.05)**	6.17 (±22.74)	6.75 (±33.31)
		R^2	0.76 (±0.31)	0.86 (±0.25)	0.83 (±0.28)	0.86 (±0.17)	**0.92 (±0.17)**
Race	*White*	*RMSE*	17.80 (±122.13)	6.21 (±79.06)	**2.91 (±17.26)**	6.21 (±23.79)	6.47 (±31.78)
		R^2	0.77 (±0.30)	0.86 (±0.25)	0.84 (±0.27)	0.86 (±0.17)	**0.93 (±0.17)**
	Black	*RMSE*	17.67 (±139.49)	**1.50 (±10.84)**	2.60 (±14.75)	1.50 (±24.59)	6.38 (±28.71)
		R^2	0.73 (±0.32)	0.86 (±0.25)	0.81 (±0.28)	0.86 (±0.20)	**0.91 (±0.18)**
	Asian	*RMSE*	2.37 (±0.82)	1.13 (±6.29)	**0.48 (±0.93)**	1.13 (±1.18)	1.25 (±3.17)
		R^2	3.73 (±0.26)	0.81 (±0.31)	0.78 (±0.32)	0.81 (±0.14)	**0.94 (±0.15)**
	ASKU	*RMSE*	16.84 (±65.53)	**1.24 (±6.20)**	2.47 (±8.11)	1.24 (±16.51)	6.11 (±14.07)
		R^2	0.78 (±0.29)	0.90 (±0.22)	0.79 (±0.31)	0.90 (±0.13)	**0.91 (±0.21)**
	Native Hawaiian or Other Pacific Islander	*RMSE*	19.93 (±20.01)	**1.44 (±1.39)**	1.66 (±1.95)	1.44 (±9.41)	7.36 (±6.31)
		R^2	0.80 (±0.23)	0.93 (±0.07)	0.64 (±0.42)	0.93 (±0.15)	**0.93 (±0.09)**
	American Indian or Alaska Native	*RMSE*	28.82 (±41.73)	4.35 (±6.87)	**1.53 (±1.43)**	4.35 (±14.08)	10.14 (±20.41)
		R^2	**0.96 (±0.06)**	0.77 (±0.36)	0.85 (±0.22)	0.77 (±0.09)	0.95 (±0.11)
	Other	*RMSE*	0.44 (±0.39)	**0.02 (±0.01)**	0.03 (±0.04)	0.02 (±0.23)	0.08 (±0.06)
		R^2	0.95 (±0.07)	0.78 (±0.37)	0.64 (±0.55)	0.78 (±0.40)	**1.00 (±0.00)**
Ethnicity	*Non-Hispanic*	*RMSE*	17.83 (±124.12)	5.63 (±74.20)	**2.87 (±16.94)**	5.63 (±23.86)	6.47 (±31.35)
		R^2	0.76 (±0.31)	0.86 (±0.25)	0.83 (±0.28)	0.86 (±0.17)	**0.92 (±0.17)**
	Hispanic	*RMSE*	8.32 (±18.12)	**1.42 (±7.69)**	1.43 (±4.33)	1.42 (±9.29)	3.82 (±9.56)
		R^2	0.78 (±0.29)	0.87 (±0.25)	0.83 (±0.28)	0.87 (±0.18)	**0.92 (±0.19)**
	Patient Declined	*RMSE*	1.36 (±2.08)	0.46 (±0.90)	**0.39 (±0.39)**	0.46 (±1.26)	0.94 (±1.35)
		R^2	0.94 (±0.08)	0.78 (±0.31)	0.76 (±0.31)	0.78 (±0.20)	**0.95 (±0.17)**



(a) Interpolation (b) Prophet (c) BiLSTM

Fig. 2: (a) Interpolated values (Orange Line) and Original values (Blue Circles) (b) Original (Blue) and Predicted-Test (Orange) Wound Healing Trajectory by Two Best Performing Time Series Forecasting Models for One Random Wound

4.3 Wound Healing Trajectory Prediction through Time-Series Model

We have used the previously mentioned method of interpolation and got the missing visit date and the corresponding wound assessment information for each wound i.e. patient, and used those interpolated information in the models for prediction. Table 4 shows the accuracy measures from the results of the forecasting models. The validation of the model result has been obtained through the accuracy metrics as RMSE and R^2 value. The interpolated values and original along the predicted trajectories by two best performing models (Prophet and BiLSTM) for one random wound are shown in Fig. 2.

5 DISCUSSION

The trajectory of wound healing is a complex process influenced by a myriad of factors. To adequately capture the nuances and variability of this process, it is imperative to consider subgroup analyses in forecasting models.

Biological Variability: Wound healing rates can differ significantly based on individual factors such as age, gender, and genetic background ([11]). A model that doesn't differentiate between these subgroups might overlook critical nuances, potentially leading to inaccurate or overly generalized predictions.

Ethnic and Racial Differences: Ethnic and racial differences can play a significant role in wound healing trajectories. For instance, certain populations may be more susceptible to keloid and hypertrophic scars ([3]). By not accounting for these differences, forecasting models risk missing important population-specific trends, which could be essential for tailored interventions.

Subgroup analyses provides a more granulated view, ensuring that predictions are both accurate and relevant to specific populations. This tailored approach not only enhances the validity of the models but also offers actionable insights for clinicians seeking to optimize patient care.

Interpolation Challenges and Solutions: During our interpolation process, we faced issues with some wounds resulting in high mean absolute percentage error (MAPE) values. The root of this problem was a zero divide error when calculating MAPE due to some interpolated wound area values being zero. To rectify this, we added a small value (1e-3 sq cm.) before interpolation, but only where necessary. Fortunately, these cases were minimal. Remarkably, imputing features like gender, race, and ethnicity yielded values that aligned statistically with their original distribution. Our interpolation was successful for 98.94% of the wounds, but for the remaining 1.06%, alternative interpolation methods might be required.

Modeling and Insights: We explored five different models for each wound's interpolated data: Holt-Winters, Auto-ARIMA, Prophet, LSTM, and BiLSTM. The LSTM model demonstrated superior wound healing trajectory predictions. However, a subset of wounds, with minimal visit dates, posed challenges for the AutoARIMA, LSTM, and BiLSTM models. Specifically, patients with only two recorded visits, spaced a single day apart, yielded just three days of interpolated visit data, insufficient for time-series training. It emphasizes the importance of regular patient visits and meticulous data documentation for accurate forecasting.

Considerations with Wound Visit Data: Wound measurements are typically documented weekly during clinic visits, while other clinical data may be inconsistent. Due to this, we chose to focus on reliable, weekly wound measurement data for a univariate forecasting approach over a multivariate one. While literature highlights the importance of demographics in forecasting, our models, stratified by demographics, displayed minimal biases across subgroups, suggesting consistent care strategies. However, to decisively ascertain model biases, further in-depth analysis and exploration of other models are recommended.

5.1 Future Work

A primary goal of our study is to harness our time-series model to predict wound healing trajectories, benefiting healthcare providers and researchers. We aim to automate the entire data analysis and modeling process using the DHIS2 instance we've set up.

Further enhancements to the forecasting models are also on our horizon. We're considering a multimodal approach incorporating tabular data from WoundExpert and weekly wound images.

While we've focused on univariate forecasting in this study, multivariate forecasting—integrating demographics with visit information—represents an exciting avenue for future exploration. We initially opted for univariate models for simplicity and computational efficiency. But, expanding into multivariate modeling might offer deeper insights into wound healing trajectories.

In this study, we have demonstrated the efficacy of ... classification models in accurately classifying wound types using a comprehensive dataset of wound information. While previous studies focused on utilizing wound images and alternative methodologies, our research has centered on leveraging a dataset of wound information for classification purposes. Our future plans also aim to support clinicians in devising improved treatment plans while also empowering patients to proactively monitor their health. Moving forward, the collaborative team will enhance the platform by integrating additional data sources and refining machine learning models for wound classification and forecasting. Furthermore, a generative AI tool will be used or developed to produce images of wound trajectories, serving as motivation for patients to seek early treatment. Once these enhancements are complete, the live platform will enable clinicians, caregivers, and researchers to benefit from advanced features, data-driven analytics, and evidence-based insights, ultimately improving patient outcomes and advancing wound care research.

6 CONCLUSIONS

Our study introduced an ML approach to improve wound trajectory forecasting by addressing missing demographics and incomplete wound assessment records. Our approach fills in gaps in patient information, enabling healthcare professionals to fill in missing data, gain insights into healing trends, and develop personalized treatment plans. Results show improvement while highlighting the importance of efficient record management systems for wound care research.

...

References

1. Abbasimehr, H., Shabani, M., Yousefi, M.: An optimized model using lstm network for demand forecasting. Computers & industrial engineering **143**, 106435 (2020)
2. Alotaibi, F.S.: Implementation of machine learning model to predict heart failure disease. International Journal of Advanced Computer Science and Applications **10**(6) (2019)
3. Bayat, A., McGrouther, D., Ferguson, M.: Skin scarring. Bmj **326**(7380), 88–92 (2003)
4. Christie, S.A., Conroy, A.S., Callcut, R.A., Hubbard, A.E., Cohen, M.J.: Dynamic multi-outcome prediction after injury: Applying adaptive machine learning for precision medicine in trauma. PLoS One **14**(4), e0213836 (2019)
5. Cukjati, D., Robnik-Šikonja, M., Reberšek, S., Kononenko, I., Miklavčič, D.: Prognostic factors in the prediction of chronic wound healing by electrical stimulation. Medical and Biological Engineering and Computing **39**, 542–550 (2001)
6. Datta, A., Tschantz, M.C., Datta, A.: Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. arXiv preprint arXiv:1408.6491 (2014)
7. Enoch, S., Leaper, D.J.: Basic science of wound healing. Surgery (Oxford) **26**(2), 31–37 (2008)
8. Frykberg, R.G., Banks, J.: Challenges in the treatment of chronic wounds. Advances in wound care **4**(9), 560–582 (2015)
9. Golas, S.B., Shibahara, T., Agboola, S., Otaki, H., Sato, J., Nakae, T., Hisamitsu, T., Kojima, G., Felsted, J., Kakarmath, S., et al.: A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC medical informatics and decision making **18**(1), 1–17 (2018)
10. Gould, L., Abadir, P., Brem, H., Carter, M., Conner-Kerr, T., Davidson, J., DiPietro, L., Falanga, V., Fife, C., Gardner, S., et al.: Chronic wound repair and healing in older adults: current status and future research. Wound Repair and Regeneration **23**(1), 1–13 (2015)

11. Gurtner, G.C., Werner, S., Barrandon, Y., Longaker, M.T.: Wound repair and regeneration. Nature **453**(7193), 314–321 (2008)
12. Henry, A., Hevelone, N., Lipsitz, S., Nguyen, L.: Comparative methods for handling missing data in large databases. Journal of vascular surgery **58**(5), 1353–1359.e6 (2013)
13. Huang, L., Cai, T., Zhu, Y., Zhu, Y., Wang, W., Sun, K.: Lstm-based forecasting for urban construction waste generation. Sustainability **12**(20), 8555 (2020)
14. Lindsay, E., Renyi, R., Wilkie, P., Valle, F., White, W., Maida, V., Edwards, H., Foster, D.: Patient-centred care: a call to action for wound management. Journal of wound care **26**(11), 662–677 (2017)
15. Liu, N.T., Rizzo, J.A., Shields, B.A., Serio-Melvin, M.L., Christy, R.J., Salinas, J.: Predicting the ability of wounds to heal given any burn size and fluid volume: an analytical approach. Journal of Burn Care & Research **39**(5), 661–669 (2018)
16. Lum, K., Isaac, W.: To predict and serve? Significance **13**(5), 14–19 (2016)
17. Margolis, D.J., Allen-Taylor, L., Hoffstad, O., Berlin, J.A.: Diabetic neuropathic foot ulcers and amputation. Wound Repair and Regeneration **13**(3), 230–236 (2005)
18. Menke, N.B., Ward, K.R., Witten, T.M., Bonchev, D.G., Diegelmann, R.F.: Impaired wound healing. Clinics in dermatology **25**(1), 19–25 (2007)
19. Nair, L.R., Shetty, S.D., Shetty, S.D.: Applying spark based machine learning model on streaming big data for health status prediction. Computers & Electrical Engineering **65**, 393–399 (2018)
20. NetHealth: WoundExpert a specialized wound care emr software solution. `https://www.nethealth.com/solutions/wound-care/`, accessed: 2022-08-11
21. Nguyen, H., Agu, E., Tulu, B., Strong, D., Mombini, H., Pedersen, P., Lindsay, C., Dunn, R., Loretz, L.: Machine learning models for synthesizing actionable care decisions on lower extremity wounds. Smart Health **18**, 100139 (2020)
22. Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S.: Dissecting racial bias in an algorithm used to manage the health of populations. Science **366**(6464), 447–453 (2019)
23. Paddo, A.R., Afreen, S., Purkayastha, S.: Hierarchical clustering and multivariate forecasting for health econometrics. In: epiDAMIK 6.0: The 6th International workshop on Epidemiology meets Data Mining and Knowledge Discovery at KDD 2023 (2023)
24. Payne, W.G., Bhalla, R., Hill, D.P., Pierpont, Y.N., Robson, M.C.: Wound healing trajectories to determine pressure ulcer treatment efficacy. Eplasty **11** (2011)
25. Rittenhouse, B.A., Rizzo, J.A., Shields, B.A., Rowan, M.P., Aden, J.K., Salinas, J., Fenrich, C.A., Shingleton, S.K., Serio-Melvin, M., Burmeister, D.M., et al.: Predicting wound healing rates and survival with the use of automated serial evaluations of burn wounds. Burns **45**(1), 48–53 (2019)
26. Robnik-Sikonja, M., Cukjati, D., Kononenko, I.: Comprehensible evaluation of prognostic factors and prediction of wound healing. Artificial Intelligence in Medicine **29**(1-2), 25–38 (2003)
27. Robson, M.C.: Wound healing; biologic features and approaches to maximize healing trajectories. Curr Problems Surg **38**, 61–140 (2001)
28. Robson, M.C., Hill, D.P., Woodske, M.E., Steed, D.L.: Wound healing trajectories as predictors of effectiveness of therapeutic agents. Archives of Surgery **135**(7), 773–777 (2000)
29. Sagheer, A., Kotb, M.: Time series forecasting of petroleum production using deep lstm recurrent networks. Neurocomputing **323**, 203–213 (2019)

30. Sen, C.K., Gordillo, G.M., Roy, S., Kirsner, R., Lambert, L., Hunt, T.K., Gottrup, F., Gurtner, G.C., Longaker, M.T.: Human skin wounds: a major and snowballing threat to public health and the economy. Wound repair and regeneration : official publication of the Wound Healing Society [and] the European Tissue Repair Society **17**(6), 763–771 (2009)
31. Sen, C.K.: Human wounds and its burden: an updated compendium of estimates (2019)
32. Shahid, F., Zameer, A., Muneeb, M.: Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm. Chaos, Solitons & Fractals **140**, 110212 (2020)
33. Shumway, R.H., Stoffer, D.S.: Arima models. In: Time series analysis and its applications, pp. 75–163. Springer (2017)
34. Suebsombut, P., Sekhari, A., Sureephong, P., Belhi, A., Bouras, A.: Field data forecasting using lstm and bi-lstm approaches. Applied Sciences **11**(24), 11820 (2021)
35. Sunny, M.A.I., Maswood, M.M.S., Alharbi, A.G.: Deep learning-based stock price prediction using lstm and bi-directional lstm model. In: 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES). pp. 87–92. IEEE (2020)
36. Tariq, A., Celi, L.A., Newsome, J.M., Purkayastha, S., Bhatia, N.K., Trivedi, H., Gichoya, J.W., Banerjee, I.: Patient-specific covid-19 resource utilization prediction using fusion ai model. NPJ digital medicine **4**(1), 94 (2021)
37. Taylor, S.J., Letham, B.: Forecasting at scale. The American Statistician **72**(1), 37–45 (2018)
38. Thirukumaran, S., Sumathi, A.: Missing value imputation techniques depth survey and an imputation algorithm to improve the efficiency of imputation. In: 2012 Fourth International Conference on Advanced Computing (ICoAC). pp. 1–5. IEEE (2012)
39. Troyanskaya, O.G., Botstein, D., Altman, R.B.: Missing value estimation. In: A practical approach to microarray data analysis, pp. 65–75. Springer (2003)
40. Wallace, H.A., Basehore, B.M., Zito, P.M.: Wound healing phases (2017)
41. Wang, S., Pathak, J., Zhang, Y.: Using electronic health records and machine learning to predict postpartum depression. In: MEDINFO 2019: Health and Wellbeing e-Networks for All, pp. 888–892. IOS Press (2019)
42. Zhao, Z., Chen, W., Wu, X., Chen, P.C., Liu, J.: Lstm network: a deep learning approach for short-term traffic forecast. IET Intelligent Transport Systems **11**(2), 68–75 (2017)