HealTech - A System for Predicting Patient Hospitalization Risk and Wound Progression in Old Patients

Supplementary Material

Subba Reddy Oota[1,2*], Vijay Rowtula[1,2*], Shahid Mohammed[1], Jeffrey Galitz[1], Minghsun Liu[1], Manish Gupta[2†]
[1]Woundtech Innovative Healthcare Solutions, [2]IIIT-Hyderabad, India
{soota, vrowtula, shmohammed, jgalitz, mliu}@woundtech.net, manish.gupta@iiit.ac.in

1. Dataset Pre-processing

Data pre-processing The following steps were taken to ensure good quality data input for our models.

- Label noise: Clinical datasets usually suffer from high inter-observer discordance. We aggregate across redundant clinician annotations to obtain a large dataset with minimal label noise.
- Occlusion: Sometimes, the wound is blocked by either scale or doctor's hand when taking the picture (refer Figure 1). So, we manually deleted some of the images where the wound part is completely blocked.
- Illumination: Wound images have been captured from a smartphone in different lighting conditions (refer Figure 1). We used grayscale normalization to circumvent the issue.
- Imbalanced data: We partially addressed this with data augmentation. Further, we experimented with over- and under-sampling to handle imbalance.
- Deformation: Similar wound images appear in different forms. So, we applied rotation and flip transformation as part of data augmentation.

2. Feature Importance Analysis

Table 1 shows the rankings of the 19 features for both the tasks: heal/hospitalization prediction and weeks to heal prediction. Age and BMI are the most important patient attributes. Wound area was the best predictor for the heal/ hospitalization classifier, which is expected since the wound area intuitively correlates with the seriousness of the wound. Similarly, wound location, area, and type are important image attributes for weeks to heal prediction.

*The first two authors made equal contribution.
†The author is also a Principal Applied Scientist at Microsoft.

Illumination Occlusion

Figure 1. Occlusion and illumination problems

No	Task 1	Task 2
1	Age	BMI
2	BMI	Wound Location
3	Wound Area	Age
4	Exudate	Wound Area
5	Adherent Yellow Slough	Wound Type
6	Pulse rate	Adherent Yellow Slough
7	Wound Margin	Exudate
8	Wound Type	Location
9	Red Granulation	Wound Volume
10	Wound Stage	Wound Margin
11	Wound Volume	Wound Stage
12	Wound Location	Pulse rate
13	Location	Red Granulation
14	Adipose Necrosis Exposed	Muscle Necrosis Exposed
15	Ligament Necrosis Exposed	Adipose Necrosis Exposed
16	Joint Necrosis Exposed	Joint Necrosis Exposed
17	Muscle Necrosis Exposed	Ligament Necrosis Exposed
18	Bone Necrosis Exposed	Bone Necrosis Exposed
19	Ethnicity	Ethnicity

Table 1. Stage 2 Results: Feature Importance Analysis (Task 1: heal/hospitalization prediction; Task 2: weeks to heal prediction). Features are listed in descending order of importance

3. Weeks to Heal Prediction Error Analysis

Figure 2 shows the error histogram for the weeks to heal prediction model. We observe that most cases have very

Wound attribute	VGG16		InceptionV3		ResNet		Xception	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Wound/Ulcer Type	0.70	0.72	0.72	0.74	0.72	0.73	**0.80**	**0.81**
Wound Location	0.82	0.80	0.83	0.83	0.86	0.86	**0.87**	**0.87**
Wound Stage	0.61	0.46	0.66	0.57	0.62	0.60	**0.63**	**0.65**
Wound Margin	0.52	0.53	**0.56**	**0.56**	0.55	0.56	**0.54**	**0.58**
Joint Necrosis Exposed	0.66	0.81	0.66	0.81	0.66	0.81	**0.79**	**0.83**
Ligament Necrosis Exposed	0.67	0.69	0.72	0.72	0.72	0.73	**0.77**	**0.70**
Adipose Necrosis Exposed	0.67	0.81	0.67	0.81	0.67	0.81	**0.69**	**0.83**
Muscle Necrosis Exposed	0.67	0.81	**0.69**	**0.83**	**0.69**	**0.83**	**0.69**	**0.83**
Exudate	0.70	0.72	**0.78**	**0.79**	0.76	0.77	0.65	0.68
Red Granulation	0.59	0.65	0.58	0.60	0.57	0.61	**0.65**	**0.68**
Bone Necrosis Exposed	0.67	0.73	0.73	0.74	0.72	0.74	**0.76**	**0.74**
Adherent Yellow Slough	0.53	0.51	0.54	0.55	0.55	0.56	**0.62**	**0.64**

Table 2. Stage 1 Results: Accuracy for wound attribute prediction using different single-task CNN classifiers. Results for Xception model are copied from "Single Task CNN" column in Table 2 in the main paper. For every attribute, we highlight results for the model with best F1.

small error, while some cases have large errors.

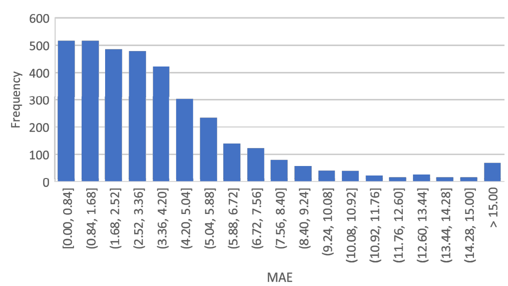

Figure 2. Error Histogram for Weeks to Heal Prediction Model

4. Weeks to Heal Data Distribution

Figure 3 shows the distribution of the weeks to heal in our dataset. As expected, the data follows a power law distribution – most wounds get healed in a very few weeks, while a very few wounds take a very long time to heal. Note that as part of pre-processing, we removed instances where the time to heal was greater than 30 weeks.

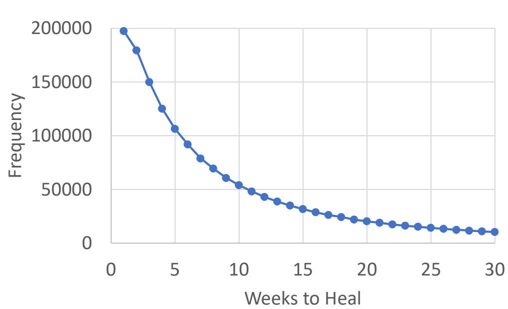

Figure 3. Error Histogram for Weeks to Heal Prediction Model

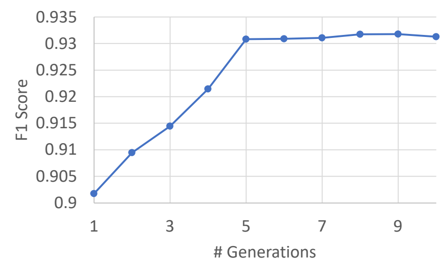

Figure 4. F1 Convergence Curve for the Evolutionary Algorithm for GS-LGBM Model

5. Comparison across multiple CNN classifiers

We tried experiments with multiple CNN classifiers like VGG16, InceptionV3 and ResNet. Results in Table 2 show that Xception models perform better for most of our tasks compared to other image prediction models.

6. Hyper-parameter tuning for Evolutionary Algorithms

We used genetic algorithm for hyper-parameter tuning to optimize the GS-LGBM model for best performance. The following features were used as chromosomes for hyper-parameter tuning: (i) number of leaves (ii) maximum depth (iii) learning rate (iv) boosting type (v) minimum child samples, and (vi) the number of estimators. The algorithm was run for 5 generations with a population size of 50 candidates per generation. The set of values (genes) per chromosome are listed in Table 3.

Fig. 4 shows variation in F1 across generations. As shown in the figure, F1 saturates after 5 generations and hence we ran the evolutionary algorithm for 5 generations for results reported in the main paper.

Chromosomes	Genes	Best Parameter
Max depth	-1, 0, 1, 2, 3, 4, 5, 6	-1
Learning rate	In the range (0.01, 0.9), with step size 0.01	0.04
n-estimators	150, 200, 250, 300, 350, 400, 450, 500	350
No. of leaves	30, 35, 40, 45, 50, 55, 60	50
Boosting type	Gradient Boosting Decision Tree, Dropouts Multiple Additive Regression Trees, Gradient based One Side Sampling, Random Forest	Gradient based One Side Sampling
Min child samples	5, 10, 15, 20, 25, 30, 35, 40	30

Table 3. Hyper-parameters for GS-LGBM Model

7. Confusion matrices for wound type, wound location and wound stage predictions.

We present detailed confusion matrices for wound type, wound location and wound stage predictions in Tables 4, 5 and 6 respectively. As mentioned in the main paper, the wound type classifier was confused between the (diabetic, pressure and surgical), and (trauma and venous) categories. As per the clinicians' diagnosis, a surgical wound at a later stage can lead to a diabetic ulcer. Similarly, a trauma wound can lead to a venous ulcer. We attribute the lower performance of our model on some of the ulcer types to the visual similarity of these ulcer types. The wound location classifier was confused between the (great toe, heel and foot), and (ankle, heel and foot) pairs. The wound stage classifier was most confused between the full thickness, partial thickness and Stage-3 classes.

8. Failure cases

We show three examples of images where the model failed to predict the correct heal/hospitalization class. Fig. 5 shows an image where actual class was hospitalization but predicted class was heal; the image has class-4 red granulation. Fig. 6 shows an image where actual class was hospitalization but predicted class was heal; the image has class-2 adherent yellow slough. Fig. 7 shows an image where actual class was heal and predicted class was hospitalization; joint necrosis is exposed in the image.

Also, in Fig. 8, we show examples of images where our wound type classifier gets confused across trauma, venous and pressure wounds. The ground truth label for these images is shown in the figure (trauma, venous and pressure). However, all these images look so similar that our classifier gets confused and predicts "venous ulcer" for all the three.

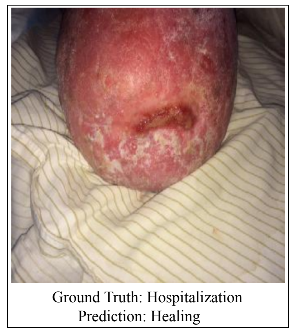

Figure 5. Error Example 1: Actual class was hospitalization but predicted class was heal

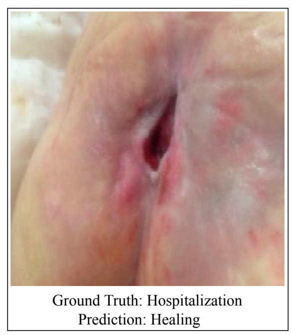

Figure 6. Error Example 2: Actual class was hospitalization but predicted class was heal

Figure 7. Error Example 3: Actual class was heal but predicted class was hospitalization

		Predicted				
		Diabetic	Pressure	Surgical	Trauma	Venous
Actual	Diabetic	3094	339	221	247	21
	Pressure	342	5883	188	228	120
	Surgical	99	124	1208	142	26
	Trauma	88	146	171	1220	241
	Venous	81	136	74	417	3160

Table 4. Confusion matrix for ulcer type classification

		Predicted					
		Sacral	Ankle	Foot	Great-Toe	Heel	Lower-Leg
Actual	Sacral	10363	1	0	8	34	126
	Ankle	38	673	78	0	49	41
	Foot	28	79	1256	51	129	20
	Great-Toe	4	1	49	157	64	12
	Heel	22	90	92	12	2036	84
	Lower-Leg	79	71	43	7	67	1651

Table 5. Confusion matrix for wound-location classification

		Predicted					
		FT	PT	Stage-2	Stage-3	Stage-4	Unstageable
Actual	FT	3977	150	19	203	34	52
	PT	283	701	11	58	9	113
	Stage-2	12	35	266	68	14	8
	Stage-3	315	143	3	1201	39	52
	Stage-4	70	15	90	112	557	3
	Unstageable	152	90	0	63	203	748

Table 6. Confusion matrix for wound-stage classification

Figure 8. Error Example 4 (for Wound Type Classifier): Actual class was Pressure, Venous, and Trauma Wounds but predicted class is venous ulcer for all the three.