Spatio-Temporal Wound Stage Classification

Alex Salman
Computer Science and Engineering
University of California, Santa Cruz
Santa Cruz, CA, USA
aalsalma@ucsc.edu

***Abstract*—With the increasing usage of neural networks and deep learning in the medical field, this work proposes an architecture for classifying wound healing stages of a series of wound images. We generate a series of consecutive wound image frames and feed them to a 2D convolutional neural network combined with long short-term memory unit and a 3D convolutional neural network to learn spatio-temporal features associated with the healing trajectory. We also visualize the saliency maps to identify features the model is extracting. Both models can extract visual features related to wound healing and have relatively high classification accuracy.**

I. Introduction

During clinical visits, wound care teams capture images of wounds and generate datasets over time. So Artificial Intelligence (AI) systems can be used for diagnosis, assessing the effectiveness of interventions, and predicting healing outcomes [1]. Research has shown the efficiencies of deep Convolutional Neural Networks (CNN) to detect and classify wound images into multiple classes [2]. The results obtained from classification methods are of high accuracy. The core to those results is the capabilities of CNN models to extract discriminatory features for predicting class probabilities [3], [4].

Interestingly, processing wound images has been revisited to use deep learning algorithms to detect wounds and estimate their size. Carrión et al. have developed a deep learning-based image analysis pipeline that aims to intake non-uniform wound images and extract relevant information such as the location of interest, wound-only image crops, and wound periphery size over-time metrics [5]. The critical indicators for healing status prediction are the rate of closure, time to closure, and expansion events. These indicators are essential to wound size and shape that change over the healing cycle.

Furthermore, in another work, Model-Agnostic Spatio-Temporal Attention Fusion (MASTAF) network [6], authors implemented feature extractions from multiple frames of images using deep neural networks. They obtained general spatial and temporal representations using three different methods, 2D-CNN, 3D-CNN, and video Transformer.

In this work, we classified frame(s) of wound into one of four classes: *hemostasis*, *inflammatory*, *proliferative*, and *maturation* (stages of wound healing) using two different methods and two different crops of wounds of the same images. The dataset used in this study is taken from a publicly available source, C57BL/6J mice dataset [7]. We used 2D-CNN together with Long Short-Term Memory (LSTM) and 3D-CNN architectures. Thus, we investigate four neural network architectures to classify two crops of wound frames. Each crop is classified on two different architectures to identify effective data modality and architecture for wound stage classification.

II. Data

The images in dataset [7] are a series of photos taken during a 15-day healing process. There are a total of 255 photos taken from four young (12-14 weeks old) and four aged (22-24 months old) mice. Each mouse received a wound on the left and right side, and photos were captured daily by cell phone at a fixed distance of 12 cm from day 0 (the surgery day) to day 15 (the experimental endpoint). One of the photos was removed due to excessive blur. In this work, we performed extracting features from two different image crops. One set of images has a red surgical splint encircling the wound, and the second is a circle cropped around the wound area itself.

Image Labels: The images were annotated by 10 non-experts using visual indicators, and the labels were aggregated using a soft voting technique. The description provided to the human annotators and the visual indicators of each wound stage (classes) are as follows:

- *Hemostasis*: Initial injury or wound infliction. Wound edges are sharp and apparent. Blood clots may be visible.
- *Inflammatory*: Begins immediately after injury. Duration varies based on the presence of infection. Redness or swelling of the wound edge might exist. The wound appears wet or shiny.
- *Proliferative*: Tissue is rebuilt through granulation tissue and re-epithelialization. Wound size is reduced by contraction. The shiny appearance of the wound does not exist as the wound dries out. The wound edge and center texture may be uneven or variable.
- *Maturation*: Tissue is reorganized and strengthened at a cellular level. The wound is healed with no open wound visible. Hair growth may be seen as skin appendages reform.

Figure 1 is an example of the four classes for the same wound of a splint crop.

III. Data Processing

We processed consecutive images into videos documenting the healing process and fed them to our neural networks to extract features. We created videos of varying lengths, from

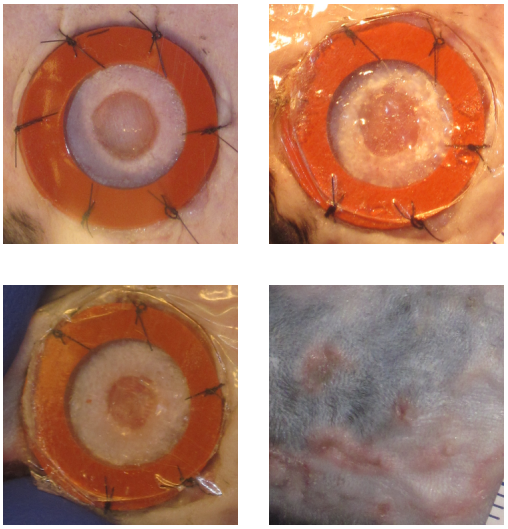

Fig. 1. Four wound classes of a splint crop: top left (hemostasis), top right (inflammatory), bottom left (proliferative), and bottom right (maturation)

a length of one frame (day) to sixteen frames (days) for each wound. Each video is padded to be 16 frames long by padding black frames to the beginning or end of wound frames to position them in the correct day index. For example, we padded two black frames on the left side of the video containing only Day 2 and Day 3 and twelve frames on the right side representing Days 4 through Day 15. For both crop versions of the data, there are 2160 videos of a length of 16 frames each.

For the dataset containing the red splints, the video frames were converted to arrays, and each frame was resized to 128×128 pixels to give an overall shape of $128 \times 128 \times 16$. Circle crop video sequences were reshaped to $96 \times 96 \times 16$. The arrays were normalized between 0 to 1. For labeling categorical data, we used Keras, to_categorical to convert a class index, integers $(0, 1, 2, 3)$ as the four classes of the wound stages to a binary class matrix. The class of the last image in the video was used as the video's wound healing stage.

Data were split into 75% train, 12.5% validation, and 12.5% test. The training data was created from 3 young and 3 aged mice and contained 1620 videos. Validation data is of the right side captures of an aged mouse and a young mouse, for a total of 270 videos. Test data is the left side captures of the same mice of the validation set, for a total of 270 videos.

See Table I for breakdown, mice numbers, and data distribution.

TABLE I
C57BL/6J MICE DATA SPLIT

Aged and Young (R+L)*	**Split Distribution**		
	Training	***Validation***	***Testing***
A8-1	R + L		
A8-3	R + L		
A8-4		R	L
A8-5	R + L		
Y8-1	R + L		
Y8-2	R + L		
Y8-3	R + L		
Y8-4		R	L
Total videos	1620	270	270

IV. MODELS AND ALGORITHMS

The pre-processed videos were sent to a 2D-CNN + LSTM and a 3D-CNN model for wound stage classification and to learn the spatio-temporal representations of each video. The following sections outline the architectures of the models.

A. 2D-CNN + LSTM

The 2D-CNN+LSTM is designed for learning spatio-temporal features from sequenced image data. Latent space embeddings created by a CNN are sent to an LSTM layer while a final dense layer generates the model output.

Splint Crop Model: We built a sequential model to take a four-dimensional input of shape $16 \times 128 \times 128 \times 3$ as D, W, H, C where C is the number of input channels. The model consists of four Conv2D layers, activation functions, MaxPooling2Ds, and Dropouts; a Flatten layer and a Dropout follow these four sets. We wrapped all the above layers with TimeDistributed wrappers to maintain frames for LSTM. A Conv2D layer requires four dimensions: $batch\_size \times height \times width \times channels$ while TimeDistributed requires an additional dimension to fit in our frames number: $batch\_size \times frames \times height \times width \times channels$. The embeddings are then forwarded to four LSTM neurons and Relu as the activation function. Lastly, the model ends with four Dense layers and a softmax layer sitting atop the LSTM.

Circle Crop Model: The architecture used to train on the 96×96 circle crop images is similar to the one for splint crop. However, we changed the flattened layer to GlobalAveragePooling2D and increased Dropout values in the fully-connected layers to avoid overfitting. Also, we added $L2$ regularizers to the convolutional and fully-connected layers.

B. 3D-CNN

The 3D-CNN architecture has fewer parameters than CNN+LSTM. Moreover, the 3D-CNN merges temporal and spatial information throughout the whole network rather than merging information with two distinct networks [8].

Splint Crop Model: The model takes in the $16 \times 128 \times 128 \times 3$ video sequences and consists of two Conv3D layers, activation functions, MaxPooling3Ds, and Dropouts; GlobalAveragePooling3D follows each dropout layer. Lastly, the two sets of convolutional layers are followed by three dense layers and a softmax activation layer.

Circle Crop Model: The architecture of this model is very similar to the one with the splint crop. The only difference is that we used a kernel size of $(1, 3, 3)$ instead of $(3, 3, 3)$ in the first Conv3D layer.

V. RESULTS

A. Accuracy and Loss

For all the experiments, each model was trained using a categorical cross-entropy loss function and Adam optimizer

with the decay of $1e-4$. Evaluation of the models was done on test data.

2D-CNN+LSTM: We trained the 2D-CNN+LSTM model on the dataset containing red splints for 150 epochs with a batch size of 54 and a learning rate of 0.0008. The overall accuracy of 66.4% was achieved. Also, we trained this architecture on the circle crops dataset for 200 epochs with a batch size of 64 and a learning rate of 0.001. Overall test accuracy of 67.7% was achieved.

3D-CNN: We trained the 3D-CNN model on the dataset containing splints for 100 epochs with a batch size of 54 and a learning rate of 0.0008. We got a test accuracy of 71.2% and a loss of 0.79. Also, we trained the model on the circle crop dataset for 500 epochs with a batch size of 64 and a learning rate of 0.001. We got a test accuracy of 73.2% and a loss of 0.66. Test accuracy of the 3D-CNN model is about 5% higher than that of the 2D-CNN+LSTM model.

Figures 2 and 3 show training and validation accuracy of the circle crop dataset on both 2D-CNN+LSTM and 3D-CNN. Additionally, all accuracy and loss values for the four models are outlines in tables II, III, IV, and V.

TABLE II
2D-CNN + LSTM MODEL TRAINED ON THE SPLINT CROP DATASET

	Data		
	Train	***Validation***	***Test***
Accuracy	92.6%	81.3%	66.4%
Loss	0.22	0.44	1.16

TABLE III
2D-CNN + LSTM MODEL TRAINED ON THE CIRCLE CROP DATASET

	Data		
	Train	***Validation***	***Test***
Accuracy	77.5%	72.8%	67.7%
Loss	0.53	0.57	1.25

TABLE IV
3D-CNN MODEL TRAINED ON THE SPLINT CROP DATASET

	Data		
	Train	***Validation***	***Test***
Accuracy	81.3%	71.7%	71.2%
Loss	0.42	0.63	0.80

TABLE V
3D-CNN MODEL TRAINED ON THE CIRCLE CROP DATASET

	Data		
	Train	***Validation***	***Test***
Accuracy	80.7%	72.4%	73.2%
Loss	0.44	0.54	0.66

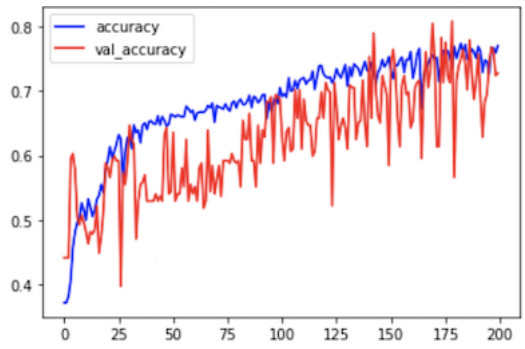

Fig. 2. 2D-CNN+LSTM Circle Crop Training and Validation Accuracy (epochs = 200)

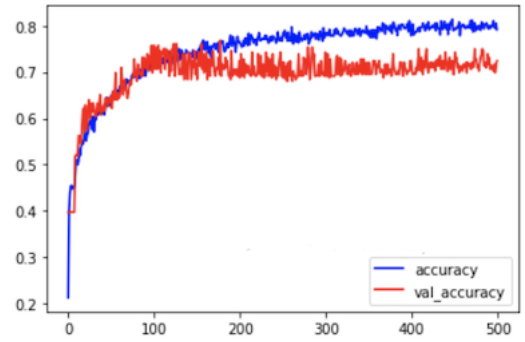

Fig. 3. 3D-CNN Circle Crop Training and Validation Accuracy (epochs = 500)

B. Visualization

In order to investigate the performance of the proposed CNN architectures in finding essential wound healing features, we extract saliency maps from the 2D-CNN + LSTM model trained on the circle crop dataset. Also, we created saliency maps for the 3D-CNN model on splint crop data to show the differences between learned activation maps for different datasets. When using the splint crop, we observed that the activation is mainly on the splint (not relying on the wound area). The focus activations are on the wound area when using the 2D-CNN + LSTM model trained on the circle crop dataset. Refer to Figures 4 and 5 to observe the same video visualized through both models.

VI. DISCUSSION

We created two deep network architectures, 2D-CNN + LSTM and 3D-CNN, for wound healing classification. For each model, two different versions were trained on two different crops of the data: images with and without splints. By training the models on the different datasets, we sought to determine which framework and image dataset results in better classification accuracy and relevant activation maps.

After conducting different training experiments for the four models, we observed that 3D-CNN performs better than 2D-CNN+LSTM and the circle crop dataset performs better than the dataset with the splint crop. This observation is also seen on the generated saliency maps. Most of the saliency maps

of the test set videos of the splint crop show highlighted activation on the splint and less on the wound. In contrast, the saliency maps of the circle crop video show different areas of the wound as activation.

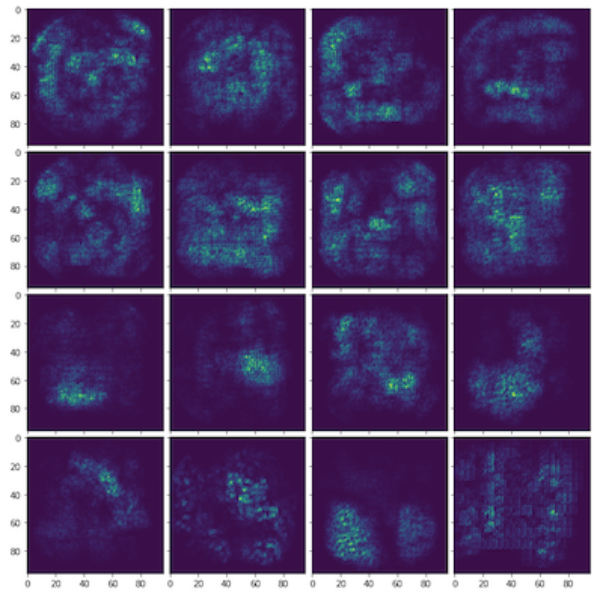

Fig. 4. Saliency maps from the 2D-CNN + LSTM model trained on the circle crop dataset.

VII. Conclusion

Using two different crops of the wound healing dataset, we generated videos from daily images of the healing process. We created four deep neural networks to classify wound stages and experimented with finding the best architecture for extracting relevant spatio-temporal features associated with wound healing. The four methods have comparable performance; however, the 3D-CNN model achieves better classification results. The saliency map visualizations further illustrate the importance of removing visual distractors from images so the model can focus on the area of interest.

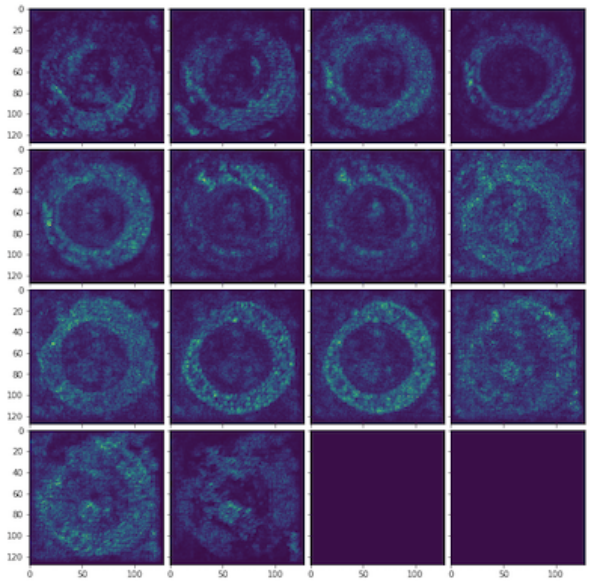

Fig. 5. Saliency maps from the 3D-CNN model trained on the splint crop dataset.

References

[1] D. Anisuzzaman, C. Wang, B. Rostami, S. Gopalakrishnan, J. Niezgoda, and Z. Yu, "Image-based artificial intelligence in wound assessment: A systematic review," Advances in Wound Care, 2021.

[2] H. Carrión, et al., "HealNet - self-supervised acute wound heal-stage classification," Medical image computing and computer assisted intervention--MICCAI 2022: 25th International Conference, 2022.

[3] B. Rostami, D. Anisuzzaman, C. Wang, S. Gopalakrishnan, J. Niezgoda, and Z. Yu, "Multiclass wound image classification using an ensemble deep cnn-based classifier," Computers in Biology and Medicine, vol. 134, p. 104536, 2021

[4] Y. Xu, K. Han, Y. Zhou, J. Wu, X. Xie, and W. Xiang, "Classification of diabetic foot ulcers using class knowledge banks," Frontiers in Bioengineering and Biotechnology, vol. 9, 2021.

[5] H. Carrión, M. Jafari, M. D. Bagood, H.-y. Yang, R. R. Isseroff, and M. Gomez, "Automatic wound detection and size estimation using deep learning algorithms," PLoS computational biology, vol. 18, no. 3, p. e1009852, 2022.

[6] R. Liu, H. Zhang, H. Pirsiavash, and X. Liu, "Mastaf: A model-agnostic spatio-temporal attention fusion network for few-shot video classification," 2021.
ye

[7] Yang, Hsin-ya; Bagood, Michelle; Carrion, Hector; Isseroff, Rivkah (2022), Photographs of 15-day wound closure progress in C57BL/6J mice, Dryad, Dataset, https://doi.org/10.25338/B84W8Q.

[8] Kwak, Geun-Ho, et al. "Combining 2D CNN and bidirectional LSTM to consider spatio-temporal features in crop classification." Korean Journal of Remote Sensing 35.5_1 (2019): 681-692.