IEEE BHI-BSN 2022
IEEE-EMBS INTERNATIONAL CONFERENCE ON BIOMEDICAL AND HEALTH INFORMATICS (BHI'22)
JOINTLY ORGANISED WITH THE IEEE-EMBS INTERNATIONAL CONFERENCE ON WEARABLE AND IMPLANTABLE BODY SENSOR NETWORKS (BSN'22)
Ioannina, Greece | September 27-30, 2022

Spatio-Temporal Wound Stage Classification

Alex Salman, Michael Briden, Narges Norouzi

Computer Science and Engineering

University of California Santa Cruz

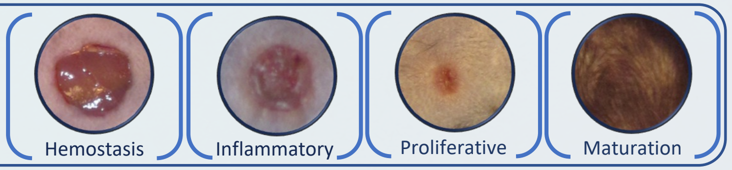
UNIVERSITY OF CALIFORNIA SANTA CRUZ

I. INTRODUCTION

The dataset:

The mice wound healing stages dataset is a series of photos taken during a 15-day healing process. The total 255 photos were 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 from day 0 (the surgery day) to day 15 (the experimental endpoint).

Figure 1 wound healing stages



The model:

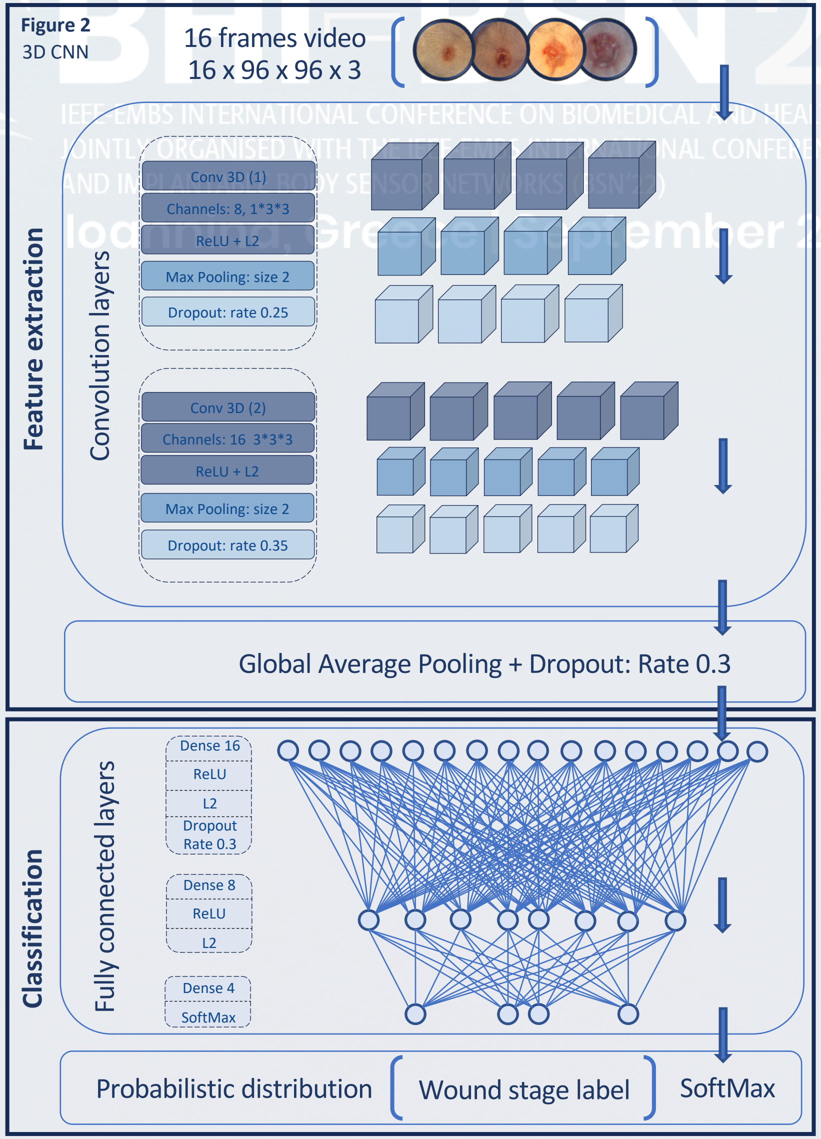
The goal of this work is to classify wound stages and find the best architecture for extracting relevant spatio-temporal features associated with wound healing. DCNNs models have shown efficiencies in wound images to extract discriminatory features for predicting class probabilities. However, in sequence classification, there are no explicit features like the one of feature vectors classification. In sequence classification, we are predicting a class label given a sequence of observations. Therefore, we need features of both space and time to predict the stages of mice wound.

Using two different datasets, splint and circle crops, we generated 2160 videos for each of the datasets of 255 images. 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 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. The 3D-CNN architecture has fewer parameters than CNN+LSTM and it merges temporal and spatial information throughout the whole network rather than merging information with two distinct networks.

II. STATE OF THE ART

Figure 2 3D CNN



This state of the art is the architecture of the 3D_CNN model. In the feature extraction part, the model takes 16 channels x 96 width x 96 height x 3 Red Green Blue video sequence input. It consists of two Convolution 3D layers, ReLU activation functions, max pooling, dropouts and average pooling. In the classification part, there are three dense layers followed by a softmax activation layer. The output is probabilistic distribution of the four wound healing stages.

III. RESULTS

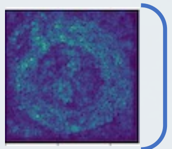
The models with no visual indicator appear to better extract desired features than the models with a surgical splint around the wound area.

Datasets	Models	Accuracy			Loss		
		Train	Valid	Test	Train	Valid	Test
Splint Crop	2D CNN + LSTM	92.6%	81.3%	66.4%	0.22	0.44	1.16
	3D CNN	81.3%	71.7%	71.2%	0.42	0.63	0.80
Circle Crop	2D CNN + LSTM	77.5%	72.8%	67.7%	0.53	0.57	1.25
	3D CNN	80.7%	72.4%	73.2%	0.44	0.54	0.66

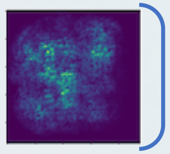
Table 1: Models Accuracy and Loss

Saliency maps showed that areas of interests are highlighted better when using no visual indicator, circle crop.

Splint Crop



Circle Crop



IV. CONCLUSIONS

The four models of the two datasets, splint and circle crops have comparable results. However, the saliency map visualizations illustrate the importance of removing visual distractors from images so the model can focus on the area of interest. These findings highlight the originality of our work by capturing the temporal features of consecutive wound frames where that was not possible in working only with individual images in previous research work.