Attention-Based CNN Model for Burn Severity Assessment

Saeka Rahman[1], Miad Faezipour[1]*, Guilherme Aramizo Ribeiro[1], Elika Ridelman[2], Justin D. Klein[2], Beth A. Angst[2], Christina M. Shanti[2] and Mo Rastgaar[1]

***Abstract*—Visual inspection, along with physical examination, is the traditional method to assess burns. However, burn-care providers have different levels of experience and may face challenges in assessing the depth and severity of the wounds. The challenges associated with the traditional approach, such as poor and varying diagnosis/prognosis accuracy, have inspired researchers towards automated burn assessment to ensure effective burn wound management. The current research aims to improve automatic burn wound assessment. It provides an ordered scoring scale to measure burn severity using four characteristics: inflammation, scar, uniformity, and pigmentation. The research also proposes an attention-based Convolutional Neural Network (CNN) model to assess the characteristics of burn wounds. The model is evaluated with 2D color images to assess levels of inflammation, scar, uniformity, and pigmentation with two different datasets, and the performances are compared with other models. The attention mechanism of the deep learning model selectively focuses on salient parts of the image to improve the understanding of the visual structure and enhance the classification accuracy. The proposed work outperforms most prior related work, achieving 93% in average accuracy.**

***Clinical relevance*—This research has significant clinical relevance in assisting accurate, reliable, and on-time diagnosis, treatment, and follow-up of burn wounds and thereby, provides effective burn wound management.**

***Index Terms*—Burn severity assessment, burn scoring scale, attention-based CNN model, channel attention, spatial attention.**

I. Introduction

Traditionally, visual inspection and physical examinations at the clinic or hospital have been the most widely used methods to assess burns in terms of severity and percentage of total body surface area (TBSA%) [1]. Burn assessment by visual inspection often leads to inaccurate or varying interpretation and estimation [2]. Pham et al. [3] systematically reviewed twenty-six relevant papers and found that the mis-estimation (of mainly burn wound size) by visual inspection ranges from 5% to 339%. Up to 77% of burns are transferred to burn centers from the referring clinics with inappropriate estimation. Such inaccurate estimation leads to inappropriate treatment, either by underestimation or resource-wasting over-estimation [4]. Doctors who are not burn-specialists can have even less accuracy in such estimation. This is while access to burn experts, especially in rural areas, may be challenging [5]. Moreover, assessing the severity of the burn wounds is more challenging than assessing TBSA% due to the mixture of depths in such wounds. Burn-care providers with limited experience face challenges in assessing the depth and severity of the wounds by visual inspection and may provide the wrong estimation, leading to inappropriate treatment and eventually inadequate healing process, including infection, scars, and degraded post-burn body function [1].

The challenges of the traditional visual inspection of burn assessment have persuaded researchers in the recent years towards automated systems that can assist in estimating TBSA% and burn severity to ensure optimal burn wound management. An automated assessment system can also provide remote assessment that can improve the initial triage and cost-effective and convenient follow-up of burn wound progression for better treatment [4], [6]. Several recent research works have been conducted to automatically assess burns using machine learning and deep learning approaches. To assess the severity, these models mostly predict burn and non-burn areas to predict the extent of burn surface (TBSA%) and depth of burns (severity).

Measuring tools like the "Lund Browder Chart", "Rule of Nines", and "Rule of Palms" are generally deployed for assessing burns [2]. These tools are subjective and provide different mechanisms to estimate TBSA%. Holm et al. [2] summarized the tool "Rule of Nine" as "body area divided into multiples of 9% body surfaces", the "Rule of Palm" as "patient's palm and fingers approximate 1% of body surfaces", and the "Lund & Browder chart" as "age-specific body areas". However, the study highlighted the issue of frequent inaccurate estimation of burns, particularly in pediatric patients and overweight patients [2].

A careful review of the literature reveals that the performance of automating burn surface area (TBSA%) assessment is improving, but not the severity assessment. The model proposed by Liu et al. [5] predicted burn and non-burn images (TBSA%) with an accuracy of 84.67%. However, the severity prediction based on the average of multiple depths was 51.44%. Karthik et al. [7] presented three models to detect and classify skin burns with accuracies of 81.4%, 80.02%, and 30.1%, respectively. Kuan et al. [1] compared 20 different machine learning algorithms with feature extraction to predict skin burn depth and found the best result of an average accuracy of 73.2%. Ribeiro et al. [8] developed a machine learning (ML)-based Linear Discriminant Analysis (LDA) model to detect scars of skin grafting in terms of inflammation, scar, uniformity, and pigmentation with accuracies of 0.86, 0.61, 0.51, and 0.80, respectively [8]. Rahman et al. [9] developed a Convolutional Neural Network (CNN) model from scratch to assess (only) inflammation of burn wounds and validated

[1]Saeka Rahman, Miad Faezipour* (corresponding author), Guilherme Aramizo Ribeiro, and Mo Rastgaar are with the School of Engineering Technology, Purdue University, West Lafayette, IN 47907, USA. rahma122@purdue.edu, mfaezipo@purdue.edu*, garamizo91@gmail.com, rastgaar@purdue.edu

[2]Elika Ridelman, Justin D. Klein, Beth A. Angst, and Christina M. Shanti are with the Department of Surgery, Division of Pediatric Surgery, Children's Hospital of Michigan/Wayne State University, Detroit, MI, USA. eridelma@dmc.org, juklein@med.wayne.edu, BAngst@dmc.org, CShanti@dmc.org

the model with 2D images. The accuracy of the model on test data was found to be 0.875. The current research explains a novel scoring scale to assess the severity of burns in terms of inflammation, scar, uniformity, and pigmentation. The research aims to improve the accuracy of assessing burn wound severity. In so doing, it proposes an attention-based CNN model to predict the inflammation, scar, uniformity, and pigmentation to assess the severity of the burns, which can improve overall burn management. The model incorporates channel attention and spatial attention within the deep learning architecture.

II. Proposed Methods

A. Burn Scoring Scale

In contrast to the subjective mechanisms of burn assessment, a novel scoring scale is introduced that provides measuring tools for burns in terms of four characteristics. These characteristics are inflammation, scar, uniformity, and pigmentation based on the color and texture of the wounds. Wounds with inflammations of different severity ‘no inflammation’, ‘mild’, ‘moderate’, and ‘severe’ are scored from 0 to 3 based on the color of the wounds which are ‘no color’, ‘pink’, ‘red’, and ‘purple’, respectively. The texture information of the wounds ‘flat’, ‘surface irregularity’, ‘raised’, and ‘hypertrophic thick’ provide scar severity in terms of values from 0 to 3. The levels of pigmentation 0 to 3 are ‘normal’, ‘hyperpigmentation’, ‘hypopigmentation’, and ‘mixed pigmentation’. Finally, texture-based information is used to define the uniformity of wounds in two levels. When the total burn area looks the same, the wound is defined as ‘uniform’ with a score of ‘0’, and when different kinds of burns are observed throughout the burn area, the wound is defined as ‘mixed’ with a score of ‘1’. The scoring scale of the four characteristics is summarized in Fig.1.

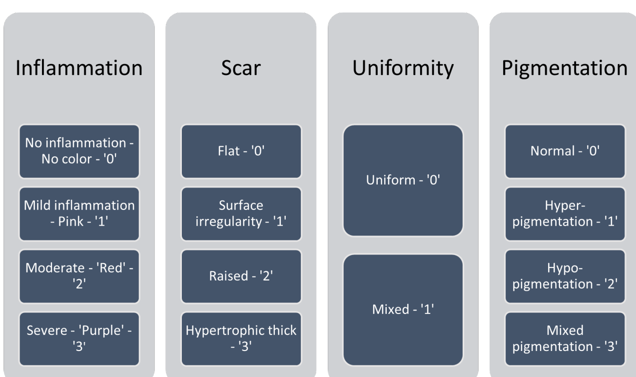

Fig. 1. Summary of the scoring scale

B. Attention-based CNN Model

CNN is a deep learning network architecture inspired by the biological brain of mammals that has demonstrated its success, especially in image classification tasks. Researchers have explored incorporating attention layers into CNNs to enhance their performance in extensive classification tasks [10], [11], [12], [13]. An attention mechanism is a valuable tool that allows systems to focus on important information while disregarding irrelevant details inspired by human attention processes [10]. When humans observe an image, they selectively concentrate on specific parts based on their needs rather than analyzing every pixel [10]. Similarly, the attention mechanism in a computer (vision) system employs selective focus on salient parts through a series of partial glimpses to improve the understanding of visual structure. Attention is divided into spatial, channel, and mixed domains, determining the areas of emphasis [10].

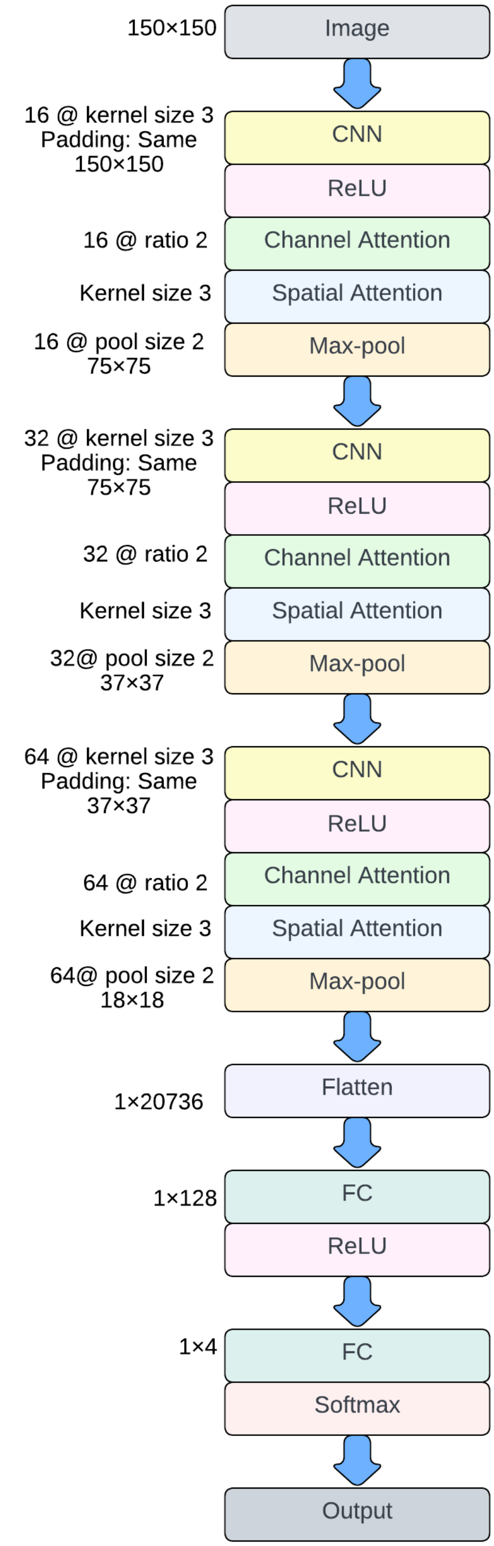

Fig. 2. Overview of the attention-based CNN model

An attention-based CNN with a suitable structure is designed for this work. The structure and layers of the proposed model is depicted in Fig. 2. The model consists of an input layer, three CNN layers, each followed by channel attention, spatial attention, and max-pooling layers, followed by fully connected (FC) layers, and finally, the classification layer. Each layer receives images from the previous layer, processes them, and feeds them to the next layer. The resized, rescaled, and augmented images are fed to the network through the input layer. Each CNN layer performs the convolution operation through a set of convolution filters and automatically extracts features from the images. The nonlinear ReLU (rectifier linear unit) activation function is applied after each convolution layer. The channel attention block performs a series of operations, including global average pooling, fully connected layers, addition, sigmoid activation, and element-wise multiplication, to generate the output emphasizing informative channels. The spatial attention block combines information from the mean and max-pooling operations and performs convolution to generate an attention-weight map that enhances feature representation and discrimination. The images are then down-sampled by max-pooling in the pooling layer. These convolution, non-linearity, channel attention, spatial attention, and max-pooling operations are repeated three times in the proposed network structure to detect features from the images. 16, 32, and 64 filters of size 3×3 are used in each convolution layer. The pool size and stride of the max-pooling layers are set as 2×2 and 2, respectively. The channel filter ratio is set to 2 and the spatial kernel size set as 3. The output from the combination of the layers is flattened to a one-dimensional vector. The layers from the flattened to the last layer are fully connected. 128 neurons are used in the first fully connected layer of the proposed model. The last layer uses the softmax activation function and classifies the class labels of inflammation, scar, uniformity, and pigmentation of burns.

III. RESULTS

Two datasets, comprised of 2D color images from different patients, are collected from the Children's Hospital of Michigan and Wayne State University, USA, in two phases. Institutional Review Board (IRB) approval for Protocol # 051717MP4X was obtained at Wayne State University for collecting and analyzing human subject data. All data has been de-identified prior to analysis. Each image is scored by burn experts using the novel burn scoring scale. Most of the pictures of the original datasets consist of multiple affected areas with different burn severity levels. Therefore, each affected region is cropped to be utilized in this research as a separate image. The subset datasets, named phase-I and phase-II, are prepared to train and evaluate the model, with descriptions presented in Table I. The prepared datasets have images from intra-patient and inter-patient division schemes where image samples of different patients are used. Phase-I dataset images are scored by a committee consensus formed by four burn care providers. Consensus scoring is not performed for the phase-II dataset. Moreover, some labels of the phase-II dataset have been combined to generate a sufficient sample size for constructing and assessing the model. To predict inflammation, the model is evaluated with 'no inflammation' and 'mild inflammation.' 'Surface irregularity' combined with 'raised' and 'flat' labels are used to assess the scar of the wounds. 'Hyper pigmentation' label, 'hypo pigmentation' and 'mixed pigmentation' are used to predict pigmentation.

The models are trained and tested by randomly dividing the dataset into 80:20 ratio. The raw images of various sizes are resized to 150×150, and the pixel intensities are rescaled to the 0 to 1 range. Horizontal flipping and random rotation are applied to augment the training dataset. The model is compiled using Adam optimizer, computing the categorical cross entropy loss, and accuracy as the performance metric. The models are trained for 150 epochs with batch size 8.

The model developed in our prior work [9] was validated using the phase-II dataset for predicting inflammation only, with an accuracy of 0.875. The current work has assessed the model [9] with the test datasets of phase-I and phase-II for all four characteristics. Moreover, we have proposed the attention-based CNN architecture in this work, which is trained and evaluated using both datasets to assess the four burn characteristics. The performance comparison of our proposed model with [9] and [8] to assess burn in terms of inflammation, scar, uniformity, and pigmentation is presented in Table II. Fig. 3 additionally presents a test instance from each characteristic group with actual and corresponding predicting classes in both datasets.

IV. CONCLUSIONS AND FUTURE APPLICATIONS

This research presents a novel scoring scale for burn assessment. It additionally implements an attention-based CNN model to predict the severity of burns from 2D color images of burn-affected areas, with a scoring scale of 0 to 3 for inflammation, scar, and pigmentation, and 0 and 1 for uniformity levels. The datasets utilized to build and evaluate the model developed in this research are labeled according to the proposed scoring scale. The model is validated to predict the four characteristic labels, and the test accuracies. The current work also validates the model presented in [9] for further assessing scar, uniformity, and pigmentation using phase-I and phase-II datasets. Finally, the performance of the attention-based CNN model is compared with the corresponding accuracies found from [8] and [9]. The proposed model is observed to outperform most prior related skin burn assessment work (e.g. [1], [5], [7]), as well as the recent ML-based LDA model developed by [8] and the CNN model presented in [9]. The proposed model can be re-tuned, and other computer vision techniques can be employed to achieve better performance. This research lays the foundation to evaluate burn wounds and create automated systems for better management of burns. Subsequent studies will also be conducted to build upon this research for more effective burn assessment and treatment. As a future direction, later steps will include finer grain assessment of the four characteristics for more complete evaluation of burns.

TABLE I
DATASET SUMMARY

Datasets	Characteristics	# of Images	Description	Labels
Phase-I	Inflammation	84	No inflammation	0
			Pink-mild inflammation	1
			Red-moderate inflammation	2
			Purple-severe inflammation	3
	Scar	75	Flat	0
			Surface irregularity-texture change	1
			Raised	2
			Hypertrophic thick	3
	Uniformity	40	Uniform	0
			Mixed	1
	Pigmentation	76	Normal	0
			Hyperpigmentation	1
			Hypopigmentation	2
			Mixed pigmentation	3
Phase-II	Inflammation	79	No inflammation	0
			Pink-mild inflammation	1
	Scar	87	Flat	0
			Surface irregularity and Raised	1
	Uniformity	89	Uniform	0
			Mixed	1
	Pigmentation	43	Hyperpigmentation	0
			Hypopigmentation and Mixed	1

TABLE II
PERFORMANCE COMPARISON

Datasets	Characteristics	Classes	LDA model [8]	CNN model [9]	**Attention-based CNN model (Proposed)**
Phase-I	Inflammation	4	0.86	0.88	**0.98**
	Scar	4	0.61	0.90	**0.94**
	Uniformity	2	0.51	0.88	**0.98**
	Pigmentation	4	0.80	0.75	**0.80**
Phase-II	Inflammation	2	–	0.875 [9]	**0.96**
	Scar	2	–	*0.88*	**0.96**
	Uniformity	2	–	*0.90*	**0.88**
	Pigmentation	2	–	*0.87*	**0.94**

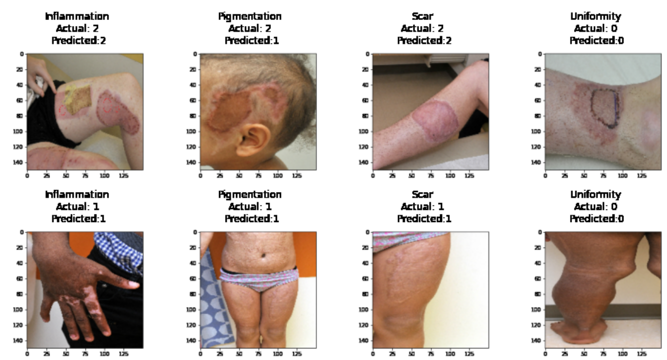

Fig. 3. Sample images with their corresponding actual and predicted labels. The top row contains the images from the phase-I dataset and the bottom row contains the images from the phase-II dataset.

REFERENCES

[1] P. Kuan, S. Chua, E. Safawi, H. Wang, and W. Tiong, "A comparative study of the classification of skin burn depth in human," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 9, no. 2-10, pp. 15–23, 2017.

[2] S. Holm, O. Engström, I. Petäjä, and F. Huss, "Does the estimation of burn extent at admission differ from the assessment at discharge?" *Scars, burns & healing*, vol. 7, p. 20595131211019403, 2021.

[3] C. Pham, Z. Collier, and J. Gillenwater, "Changing the way we think about burn size estimation," *Journal of Burn Care & Research*, vol. 40, no. 1, pp. 1–11, 2019.

[4] D. Parvizi, M. Giretzlehner, J. Dirnberger, R. Owen, H. Haller, M. Schintler, P. Wurzer, D. Lumenta, and L. Kamolz, "The use of telemedicine in burn care: development of a mobile system for TBSA documentation and remote assessment," *Annals of burns and fire disasters*, vol. 27, no. 2, p. 94, 2014.

[5] H. Liu, K. Yue, S. Cheng, W. Li, and Z. Fu, "A framework for automatic burn image segmentation and burn depth diagnosis using deep learning," *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021.

[6] B. Z. Barkana, D. A. Barkana, and M. Faezipour, "Image-based determination of the growth or shrinkage of wounds at the dermal layer," in *International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2020, pp. 1706–1707.

[7] J. Karthik, G. S. Nath, and A. Veena, "Deep learning-based approach for skin burn detection with multi-level classification," in *Advances in Computing and Network Communications*. Springer, 2021, pp. 31–40.

[8] G. A. Ribeiro, E. Ridelman, J. D. Klein, B. A. Angst, C. M. Shanti, and M. Rastgaar, "50 assessment of skin graft in pediatric burn patients using machine learning is comparable to human expert performance," *Journal of Burn Care & Research*, vol. 41, pp. S33–S34, 2020.

[9] S. Rahman, M. Faezipour, G. Ribeiro, E. Ridelman, J. Klein, B. Angst, C. Shanti, and M. Rastgaar, "Inflammation assessment of burn wound with deep learning," in *International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2022, pp. 1800–1803.

[10] W. Xiao, H. Liu, Z. Ma, and W. Chen, "Attention-based deep neural network for driver behavior recognition," *Future Generation Computer Systems*, vol. 132, pp. 152–161, 2022.

[11] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5659–5667.

[12] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[13] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.