University of Wisconsin Milwaukee

UWM Digital Commons

May 2021

Wound Image Classification Using Deep Convolutional Neural Networks

Behrouz Rostami
University of Wisconsin-Milwaukee

Follow this and additional works at: https://dc.uwm.edu/etd

Part of the Computer Sciences Commons, and the Electrical and Electronics Commons

Wound Image Classification using Deep Convolutional Neural Networks

by

Behrouz Rostami

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Engineering

at

The University of Wisconsin–Milwaukee

May 2021

ABSTRACT

Wound Image Classification using Deep Convolutional Neural Networks

by

Behrouz Rostami

The University of Wisconsin–Milwaukee, 2021
Under the Supervision of Professor Zeyun Yu

Artificial Intelligence (AI) includes subfields like Machine Learning (ML) and Deep Learning (DL) and discusses intelligent systems that mimic human behaviors. ML has been used in a wide range of fields. Particularly in the healthcare domain, medical images often need to be carefully processed via such operations as classification and segmentation. Unlike traditional ML methods, DL algorithms are based on deep neural networks that are trained on a large amount of labeled data to extract features without human intervention. DL algorithms have become popular and powerful in classifying and segmenting medical images in recent years. In this thesis, we shall study the image classification problem in smartphone wound images using deep learning. Specifically, we apply deep convolutional neural networks (DCNN) on wound images to classify them into mutiple types including diabetic, pressure, venous, and surgical. Also, we use DCNNs for wound tissue classification. First, an extensive review of existing DL-based methods in wound image classification is conducted and comprehensive taxonomies are provided for the reviewed studies. Then, we use a DCNN for binary and 3-class classification of burn wound images. The accuracy was

considerably improved for the binary case in comparison with previous work in the literature. In addition, we propose an ensemble DCNN-based classifier for image-wise wound classification. We train and test our model on a new valuable set of wound images from different types that are kindly shared by the AZH Wound and Vascular Center in Milwaukee. The dataset has been shared for researchers in the field. Our proposed classifier outperforms the common DCNNs in classification accuracy on our own dataset. Also, it was evaluated on a public wound image dataset. The results showed that the proposed method can be used for wound image classification tasks or other similar applications. Finally, experiments are conducted on a dataset including different tissue types such as slough, granulation, callous, etc., annotated by the wound specialists from AZH Center to classify the wound pixels into different classes. The preliminary results of tissue classification experiments using DCNNs along with the future directions have been provided.

To my love,

Elaheh

Table of Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to thank my mentor and research advisor, Prof. Zeyun Yu, for all of his continuous supports, invaluable advice, and patience during my PhD studies. Your guidance and encouragement besides knowledge and constructive comments have always helped me to deal with challenges and overcome obstacles. Thank you for providing me with the opportunity to work under your supervision.

I would like to express my sincere gratitude to my thesis committee members: Professor Hu, Professor Dsouza, Professor Zhang, and Professor Wang for their insightful and important comments.

My sincere thanks also goes to Dr. Gopalakrishnan for all his help and supports and my labmates for their assistance and kindness.

Nobody has been more important to me in the pursuit of this degree than my family members. I can't thank my dear parents enough who set me off on the road to this PhD a long time ago, and whose love, invaluable advice, encouragement, guidance, and supports are with me always and in whole my life. From the bottom of my heart I would like to thank Reihaneh, Ramin, and Farid for all your helps, counseling, and encouragement over the years.

I wish to thank my lovely and supportive wife, Elaheh, for helping me throughout this challenging career. You made me stronger in deal with obstacles and challenges and I couldn't have done it without you. Thank you for everything, my love.

Chapter 1

Introduction

1.1 AI and Machine Learning in Healthcare

Artificial Intelligence (AI) is a field of computer science which discusses intelligent systems that mimic human behavior and includes a wide vriety of subfields and branches [2]. In recent years, AI algorithms have evolved into so-called data-driven techniques without human or expert intervention as opposed to the early generations of AI that were rule-based relying largely on an expert's knowledge [3]. As a potential solution for a wide range of human life problems, AI and its subbranches like Machine Learning (ML) have been used in different areas from banking and finance to manufacturing, policing, and marketing [4, 5]. Specifically, healthcare can be named as an important field in which AI and ML footprint can be found frequently [3]. Cancer, neurology, cardiology, radiology, ophthalmology, immunology, genetics, and wound care are just a few examples in healthcare that benefited from AI and ML in the recent years [6, 4, 7, 8, 9, 10, 11, 3]. Moreover, recent advances in computer vision techniques including object classification, detection, and segmentation has been resulted in using ML widely in some healthcare subfields such as medical imaging [12]. Also, ML has enabled many image analysis applications of Computer-Aided Detection (CAD) systems. Such systems have shown remarkable

results when applied to different fields including healthcare [13]. CAD systems built for medical image analysis, assist medical care providers and radiologists with decision making to improve the detection of abnormalities by reducing the false negative rate [14, 15]. Lung nodule detection [16], skin cancer classification [17, 18], Diabetic retinopathy detection [19, 20], and breast cancer analysis are a few out of many CAD systems from the literature.

Traditional machine learning algorithms are mostly knowledge-driven techniques in which an expert should define ad-hoc features to be used for training the algorithm. On the other hand, Deep learning (DL) methods which are a subset of ML, are trained to extract hierarchy of features out of large volumes of data usually without any need for extensive image pre-processing or hand-crafting features (i. e., data-driven techniques) [21]. In fact, by using several hidden layers, deep networks are able to capture and model very complex relationships between the system's input and output [22]. Since 2012 that a deep learning model, AlexNet, showed a superior performance in image analysis over traditional machine learning techniques, several deep learning approaches proposed for data analysis in different fields [2]. Furthermore, Transfer Learning (TL) accelerates model training significantly as well as improving the model's performance [23, 24]. In fact, TL is a technique that enables researchers to apply a knowledge which was learned on a different domain into a new problem [25]. It became more popular in deep learning algorithms as training these networks require tremendous amount of data. Therefore, utilizing a pre-trained model can surmount the difficulties of training models from scratch resulting in a more efficient model training. This is critical for the medical image analysis appli-

2

cations due to the data leakage problem as a usual issue in the field [26].

1.2   Wound Management - Medical Background

As one of the most important organs in human body, skin has several vital roles including protecting the internal body organs from external harmful factors [27]. If this important organ is damaged, its functions will be affected significantly [27]. Considering the origin for this damage, wounds are classified under two main categories: acute or chronic. In fact, acute group includes those wounds which are caused by an external agent, while chronic wounds are originated from an internal cause [28]. The external cause includes agents like bites, burns, and minor cuts [28]. Also, agents such as neoplasm, impaired venous drainage, and some metabolic deseases including diabetes can be named as internal causes [28]. Each of acute and chronic category includes different types of wound. For example, acute group include types like surgical and burn wounds and chronic group includes types such as venous, arterial, pressure, and diabetic [28].

Acute and chronic wounds are a challenge and burden to healthcare systems worldwide. In the United States alone, acute wounds affect 11 million people and chronic wounds influence more than 6 million humans annually with an estimated medicare burden of \$28.1 billion to \$96.8 billion (US) [29, 30, 31]. Also, in Europe, based on estimations, 1.5 to 2 million people are affected by acute or chronic wounds [32]. Moreover, it has been estimated that there are nearly 3500 people in a population of 1 million, will experience having a wound [32]. From this affected population, for

525 people it takes more than 1 year for the wound to be healed [32]. In addition, estimations show that 14-24% of diabetic patients that developed foot ulcers will experience an amputation [33]. Specifically, it has been shown that chronic wounds have many mental effects including patient's depression and social separation in addition to physical problems [34]. This information, in addition to other risks like wound infection show that how having an efficient wound management plan is critical for both patients and physicians and may result in a better wound treatment and healing process. Early diagnosis is an important step in wound detection and may result in having faster healing process [35]. Visual investigations and evaluations are currently the most common approaches for wound assessment and analysis. However, these methods are prone to human errors [35]. In the recent years, many researchers used AI and its subbranches to solve the mentioned issues. Next subsection discuss advances in one of the most common areas in wound management, wound image analysis.

1.3   Wound Image Analysis

The characterization of a wound is a key step in wound diagnosis that would help clinicians to identify an optimal treatment procedure and assess the efficacy of the treatment [36]. Many studies proposed wound image-based analysis approaches to make the wound diagnosis more efficient. Wound image analysis includes different tasks such as classification, segmentation, and tissue analysis. Wound classification is an important step of wound image anlysis and defines as classifying a wound as

a whole into different types (e.g., venous, diabetic, pressure) or different conditions (e.g., ischemia vs. non-ischemia, infection vs. non-infection) and constitutes an essential part of wound assessment [37]. Moreover, in some cases visual analysis of the wound is the approach usually used for wound analysis which is not accurate and may result in human errors [38]. Therefore, having an efficient wound classifier with a reliable classification performance will result in saving time and money [36]. Also, having a tool that is able to specify and separate different wound tissue types automatically, helps the wound specialists to have better wound healing monitoring and more precise measurments [39].

Many papers have been published in the wound care field within the recent years. Some of them discussed the wound classification problem using a traditional machine learning algorithm [40, 41] and part of the researches studied the wound classification problem using more modern techniques like DCNN-based methods [42, 43, 44]. However, only a few papers were identified that discussed wound analysis from the wound type classification point of view. Also, most of the publications on wound type classification topic, discussed only the binary classification problems such as classifying the samples into normal and abnormal cases. Having difficulties to access a reliable dataset can be mentioned as a reason for this issue.

In the wound analysis part of this thesis we will use deep convolutional neural networks for wound image analysis including wound type and tissue classification. The first phase of our wound image analysis project is wound type classification. Initially, we do the classification for burn wound images and then study the classification of chronic wounds. In more detail, by collaborating with a well known wound

clinic in Milwaukee, AZH Wound and Vascular Center, we obtained a valuable set of wound images from different types including diabetic, venous, pressure, and surgical. We propose a DCNN-based classifier to classify the wound images into two or multiple classes. The second phase would be the tissue type classification or semantic segmentation of the wound tissues. In this step, we use DCNNs for classifying each pixel of the input wound image into one of the predefined tissue types including granulation, necrotic, slough, callous, and so on.

1.4 Contribution and organization

In this thesis, (1) we review the existing deep learning-based acute and chronic wound image classification studies extensively and suggest a comprehensive taxonomy of existing works in burn and chronic wound classification. To the best of our knowledge these are the first taxonomy in the literature. (2) We use a deep convolutional neural network for classifying burn wound images into 2 and 3 classes. We test our classifier using a bublicaly available burn wound image dataset and compare our results with one of the previous works in the literature. Our results showed a considerable improvement on the classification accuracy for the binary classification problem. (3) We introduce a new dataset of real wound images, collected by AZH Wound and Vascular Center, containing image data with more wound types than those considered in the prior publications including diabetic, pressure, venous, and surgical wound types. All these images along with their types and corresponding extracted ROIs will be made publicly available. (4) Numerous experiments were

6

conducted for patch-wise wound image classification and the prformance of different patch classifiers were evaluated to be used as a building block of a whole image wound classifier. (5) As the main contribution of this thesis, we propose an end-to-end ensemble DCNN-based classifier to classify the entire wound images. To the best of our knowledge, this research is the first study in which an ensemble deep learning-based method used for image-wise classification of the wound images into multiple types. We have proposed an innovative combining strategy to combine a patch classification-based classifier and a common DCNN, AlexNet to classify the wound images. Transfer learning technique is utilized and the pretrained deep convolutional neural networks are fine-tuned by our own wound image dataset to provide a more efficient classification performance. By accepting the entire wound image as the input for the model, our proposed approach is able to generate the wound type as the output. (6) The performance of the proposed model is evaluated by testing on our own wound image dataset as well as comparing with some common deep convolutional neural networks used in the literature and testing on a publically available wound image dataset. (7) We review comprehensively the recent articles in the filed wound tissue classification and analysis and provide an extensive taxonomy for the reviewed suties. To the best of our knowledge, this is the first taxonomy in the literature. (8) We introduce a new wound image dataset including different tissue types annotated by the AZH Center's specialists. The dataset will be made publically available after collecting more images and getting the required permissions from the clinic. (9) We use deep convolutional neural networks along with SLIC superpixel generation method for classifying the wound image pixels into different tissue types

7

including slough, granulation, callous, etc.

The rest of this thesis is organized as follows: In chapter 2, an introduction to deep learning is provided. Also, we discuss some of the most common deep architectures that were used in the literature in detail. At the end of this chapter, we discuss some applications of deep learning in healthcare and wound care. In chapter 3, first we discuss the burn wound image classification topic which includes reviewing the previous works, the utilized dataset, our method, and the results. Chapter 4 is dedicated to the multiclass wound image classification topic which is the core part of this thesis. This chapter includes a comprehensive review on the recent studies in the literature, as well as talking about our dataset, the proposed method, and the results. Chapter 5 discusses the pixel-level classification of wound images which is the next step in wound image classification task. Finally, this thesis included in Chapter 6 along with talking about the future directions.

Chapter 2

Deep Learning

2.1 Introduction

In its early stages, Artificial Intelligence (AI) was used to solve the problems that were hard for human but easy for computers. An example for these problems is defeating a chess champion by a computer that was happened in 1997 [45]. Gradually, humans started to use AI to do tasks that are hard for computers, but easy for people. In fact, people can do these tasks easily, but explaning them formally and by using a set of rules is hard. In order to use computers for solving the mentioned problems, human needed to find a way to transfer knowledge to them [45]. In Machine learning, computer can extract these knowledge from the raw data. By finding the relation between some manually extracted features and the system's output, computer is be able to detect the outcome for a new input. The fact that how we define the features to the system has an important role in its performance. Also, there are many problems in which knowing about what features are the best for that specific task, is difficult [45]. In representation learning, the feature extraction step is done automatically which is resulted in saving time. Also, the system is able to define the best features by its own. However, simple or classic machine learning tools are not able to extract very high-level features that are needed for more complex tasks [45].

9

Deep learning has solved this issue by extracting different levels of features and constructing the higher level complex features on top of the early stage simpler features. Deep learning models are advanced version of simple Multilayer Perception (MLP) [45]. In spite of their anticedents that had only a few hidden layers, deep networks include several sequential hidden layers that make it possible to extract more informative features [45, 6]. In fact, deep learning is a representative of the data-driven developments, which is able to analyze complicated data automatically and extract the needed information, relationships, and patterns [46, 4].

2.2 Deep Convolutional Neural Networks (DCNN)

Deep convolutional neural networks are one of the most common deep learning models with a varying number of hidden layers. They are different from the classical neural networks that have a limited number of layers between the network's input and output [6]. In a DCNN, convolution is the main mathematical operation to process the network's input [21]. In more detail, Convolutional Networks (ConvNet) are usually built from layers of convolution, sub-sampling, and optional fully connected layers [47]. Unique characteristics of ConvNets such as weight sharing and transition invariance that result in reduced number of parameters and simplify calculations have transformed them to one of the most popular deep learning architectures [21]. In the rest of this chapter, we introduce some of the most famous DCNNs and discuss their structure in more detail. Most of the reviewed deep architectures in this section are winner algorithms of ImageNet Large Scale Visual Recognition Challenge

(ILSVRC) [48]. ILSVRC is an annual competition which evaluates the object detection and image classification performance of different methods in large scale. By holding the competition researchers can compare detection methods in a wider range of objects and measure the advancement of machine vision based approaches in large scale image tasks such as annotation. We refer the interested readers to a very good survey on deep learning approaches [49] to get a more detailed summary of each of the described networks.

2.2.1 AlexNet

AlexNet is a deep CNN architecture proposed in 2012 as the winner of ILSVRC. It was able to beat all of the traditional machine learning methods [1, 49]. This network has 8 layers with 60 million parameters. The first five layers of the network are convolutional layers and the last three layers are fully-connected layers. The last fully connected layer is connected to a softmax that generates 1000 probability values for 1000 class labels [1]. Figure 2.1 shows the AlexNet architecture.
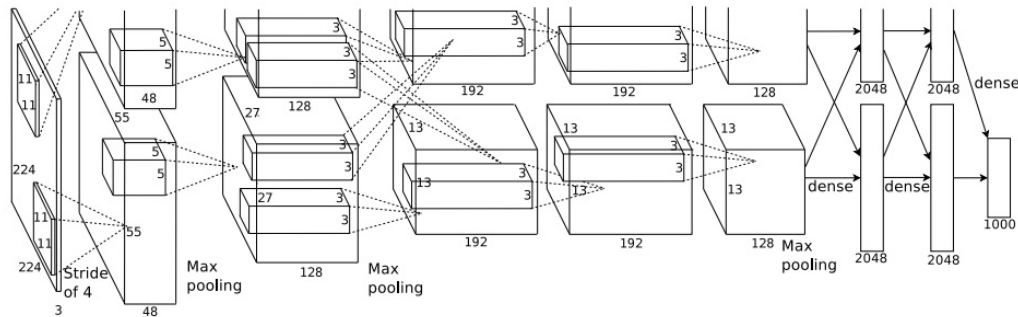
Figure 2.1: AlexNet architecture [1].

2.2.2 VGGNet

This deep CNN which proposed in 2014 by Visual Geomtry Group, indicated that by increasing the netwrok's depth, higher accuracy values can be achieved in some task such as classification [50]. Three VGGNet models proposed with the names VGG-11, VGG-16, and VGG-19 which the numbers at the end of the names shows the number of layers in the network. VGG-19 has 138 million parameters [49].

2.2.3 GoogLeNet

GoogLeNet is another deep convolutional neural network proposed in 2014 [51]. This network is a particular shape of the Inception architecture and includes 22 layers. The important feature of this network is dimension reduction. Comparing with AlexNet and VGGNet, the number of parameters are considerably lower in GoogLeNet which reduces the complexity of calculations [49]. Specifically, in comparison with the AlexNet architecture proposed in 2012 [1], GoogLeNet provides 12 times fewer parameters [51].

2.2.4 ResNet

This residual learning-based deep convolutional neural network proposed in 2015 in different versions such as ResNet18, ResNet34, ResNet101, ResNet152 which have different number of layers [52]. ResNet50 is one of the must popular ResNet models with 50 convolution layers and 25.5 million parameters [49]. Despite of being deeper in comparison with some networks like VGGs, ResNet has lower complexity [52]. In

2015 an ensemble of ResNet architectures outperformed the other methods on the ILSVRC classification task and won the first place [52].

2.2.5 InceptionV3

InceptionV3 is a deep architecture proposed in 2014 as one of the versions of the Inception network [53]. There are different versions for the Inception network such as InceptionV1 , InceptionV2, and InceptionV3 [53]. Having 42 layers, InceptionV3 network utilized factorization concept to decrease the number of parameters of the networks and computational complexity. Also, InceptionV3 is faster than its counterparts because of its parallel network implementation [54].

2.3 Deep learning in healthcare

Deep Learning has been utilized widely in various fields including healthcare and generated considerable outcomes, especially in medical image analysis domain. In fact, the increasing growth of the data complexity and volume in the medical imaging realm has resulted in broadly utilization of deep learning in areas like image classification, segmentation, and detection [6, 2]. Detection of organs and body parts in MR or CT images, cell detection in histopathological images, and computer-aided detection and diagnosis are health care examples in which researchers used DL and its derivatives such as Deep Convolutional Neural Networks (DCNN)[55]. However, one of the challenges in using deep learning in healthcare is data limitation problem. Most of the deep learning models need a huge number of data to be trained well and

for some clinical applications it is not easy to provide this volume of data [56]. In addition, being not interpretable is another challenge about DCNNs. In fact, DCNNs are black boxes and there is not the possibility to explain why they are generating good or bad results [57].

2.4   Deep learning in wound care

At prersent, physicians perform visuall investigations and analysis to obtain some wound measurements such as area and depth [36, 43]. Many studies have been conducted by reaserchers in the wound care field recently in which they proposed automated DL [58, 59, 60] and DCNN-based approches for wound assessment. These studies cover a wide range of wound image analysis tasks like segmentation [61, 38] and classification [62, 63, 64]. Diabetic foot ulcer classification, classification of burns and pressure ulcer, and dermatological image classification are examples of these studies in wound classification domain. Also, there are some studies in the literature in which the authors used DCNNs for wound tissue classification [65, 66, 67]. In these studies, the goal is to classify the wound pixels in the image into different tissue types like granulation, fibrin, necrosis, etc. [65]. All of the mentioned studies, have clearly shown the effectiveness and efficiency of deep convolutional neural networks in wound diagnosis and analysis.

2.5 Performance Metrics

In this subsection we provide the equations we used in this thesis for investigating the performance of the classifiers. The performance metrics formulae have been provided below. More derails about the equations 2.1 to 2.4 can be found in [68]. In addition to these metrics, we used the Area Under the ROC Curve (AUROC or AUC) metric in binary classification problems.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2.2}$$

$$Recall(TPR) = \frac{TP}{TP + FN} \tag{2.3}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2.4}$$

$$Loss = -\sum_{i=1}^{N}\sum_{j=1}^{K}(t_{i,j}\ ln\ y_{i,j}) \qquad (2.5)$$

In these equations, TP, TN, FP and FN represent True Positive, True Negative, False Positive, and False Negative measures respectively. Also, the equation 2.5 shows the cross entropy formula used for calculating the Loss function value in classifiers in this thesis [69].

Chapter 3

Burn Wound Image Classification

In this chapter we discuss wound image classification using deep convolutional neural networks. In wound classification task, the goal is to classify the input wound image into two or multiple classes. This chapter has been organized under two main subsections: burn wound image classification and multiclass wound image classification. In each subsection, we discuss the previous works in the literature as well as implemented models and their corresponding results. The chapter is concluded in Section 4.6.

3.1 Related works

In this subsection, some of the recent publications in burn wound image classification have been reviewed and categorized under two main subbranches: feature extraction methods along with an SVM, and end-to-end DCNN-based approaches. Figure 3.1 displays the complete organization chart for the studied papers.

3.1.1 End-to-end DCNN-based approaches

Chauhan et al. used deep learning approaches to classify burn images based on the part of the body recognized in the image [70]. The proposed method used ResNet-
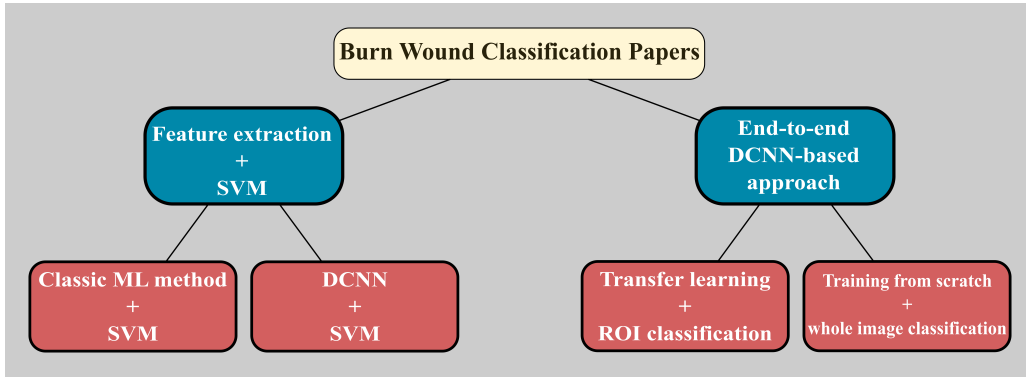
Figure 3.1: Organization chart for the burn wound classification papers.

50 architecture to categorize the input burn images into the face, hand, back, and inner arm classes. The utilized dataset included 109 burn images collected from the web as well as 4981 non-burnt images obtained from some available datasets. Two dependent (which used leave-one-out cross-validation) and independent (which used an independent deep learning model) approaches were tested in this research. The second strategy generated a better classification performance with an accuracy value of higher than 93%.

In another article, Cirillo et al. applied deep convolutional neural networks on burn images to predict the wound depth [71]. The goal was to classify the burn images into four classes: deep partial-thickness and full-thickness depth, intermediate to deep partial thickness, superficial to intermediate partial thickness, and the superficial partial thickness. A total of 23 burn images collected by a hospital in Sweden were used for training the models. 676 extracted ROIs from six classes (four burn-depth classes as well as normal skin and background categories) were augmented to overcome the overfitting problem. Four pre-trained deep architectures includ-

18

ing VGG-16, GoogleNet, ResNet-50, and ResNet-101 were tested as classifiers. The ResNet-101 generated the best classification performance with the average classification accuracy of 91%. Moreover, the authors claimed that the data augmentation improved the classification accuracy.

3.1.2 Feature extraction + SVM

Yadav et al. suggested a machine learning-based approach for binary classification of burn wound images [40]. A traditional color-based feature extraction method was used in company with a support vector machine to classify the images into two classes (grafted and non-graft). The utilized dataset contained 94 images from three burn types including full-thickness, deep dermal, and superficial dermal. The first two types constituted the "grafted" class while the "non-graft" category consisted of the superficial dermal images. The reported classification accuracy was 82.43%.

Abubakar et al. used DCNNs for binary classification of pressure and burn wound images [72]. Different deep architectures including VGG-face, ResNet101, and ResNet152 were applied for feature extraction followed by an SVM to classify the images into burn or pressure categories. The dataset was collected from the internet as well as a hospital source and contained 29 pressure and 31 burn wound images. Cropping, rotation, and flipping transformations were utilized for data augmentation. Also, several binary (burn or pressure) and 3-class (burn, pressure, or normal skin) classification experiments were conducted. For both classification problems, the best performance was related to ResNet152 which resulted in classification accuracy of 99.9%.In a second study, Abubakar et al. proposed a classification method based on

deep learning to classify burn wound images into Caucasian and African patient categories [73]. Three pre-trained deep convolutional neural networks including VGG-16, VGG-19, and VGG-Face, were used for feature extraction from the input images. In the next step, an SVM classified the extracted features into one of the healthy or burn classes. For each of the deep architectures, three different datasets including African patients, Caucasian patients, and a combination of them were utilized to train the SVM. The intent was to study how the combination of the images from different skin colors affects the classifier's performance. The dataset consisted of 32 Caucasian and 60 African cases. During the patch generation step, 1360 and 700 patches were extracted from the two groups, respectively. The authors mentioned that the classification accuracy was higher for Caucasian patients in comparison to the African patients or to the hybrid group. The combination of VGG-16 and SVM was reported as the best classification strategy with an accuracy of 99.286% for Caucasian patients, 98.869% for African patients, and 98.750% for the hybrid dataset. ResNet101 architecture followed by an SVM classifier was used in another research by Abubakar et al. to classify burn wound images into one of the two classes, burn or normal [62]. A pre-trained ResNet101 architecture was utilized as a feature extractor and then an SVM was applied for classification. The dataset was collected from the internet and after augmentation included 1360 images. The reported accuracy and precision values were 99.49% and 99.56%, respectively. Also, the outcomes were compared with LeNet's performance in a similar article which presented 81.81% for the precision metric.

Table 3.1 summarizes the reviewed papers. Based on this literature review, only

a few articles studied the burn wound image classification problem and most of them only discussed the binary classification task. Additionally, it is important to note that only a limited number of researches used an automatic end-to-end deep learning-based method for classification. Instead, at the final stage they utilized a traditional ML tool as the classifier. To address this research gap, we propose an end-to-end DCNN-based approach to perform binary and 3-class classification of burn wound images.

3.2   BIP_US dataset

In this research, the same dataset that was utilized in a prior investigation [40] was used for comparing our results with those previously reported. This dataset, BIP_US, contains 94 images from three burn wound types including full-thickness, deep dermal, and superficial dermal [74]. There are 20, 32, and 42 samples in each class, respectively. The images have jpg and bmp formats and they are in different sizes. Figure 5.2 shows some sample images from the dataset.

3.3   Methods

3.3.1   Preprocessing

Data splitting:

In the 3-class classification problem we split the data into train, validation, and test sets with 76, 9, and 9 images, respectively. For the binary classification experiment,

Table 3.1: Summary of burn wound image classification works.

Work/Research	Classification	Methods	Dataset
Yadav et al. [40]	Binary classification of burn images into wounds that needs graft and the non-graft wounds	Support Vector Machine (SVM).	Burns BIP_US Database.
Abubakar et al. [72]	Binary Classification of wound images into burn wounds or pressure wounds.	VGG-face, ResNet101, and ResNet15 networks with SVM classifier	A dataset with 29 pressure wound images and 31 burn wound images.
Abubakar et al. [73]	Binary Classification of input images into healthy or burn wounds.	VGG-16, VGG-19, and VGG-Face with SVM classifier	A dataset with 32 wound images from Caucasians and 60 wound images from Africans
Abubakar et al. [62].	Binary Classification of input images into burn skin or normal skin.	ResNet101 architecture and SVM classifier	An unknown number of burn wound images collected from the internet
Chauhan et al. [70].	Classifying the body part of burn images into four classes: back, face, hand, and inner forearm	ResNet-50 architecture	109 burn images collected from the web as well as 4981 non-burnt image
Cirillo et al. [71].	Classifying the burn images into four classes: deep partial-thickness and full-thickness, intermediate to deep partial thickness, superficial to intermediate partial thickness, and superficial partial thickness	VGG-16, GoogleNet, ResNet-50, and ResNet-101	23 burn images collected by a hospital in Sweden
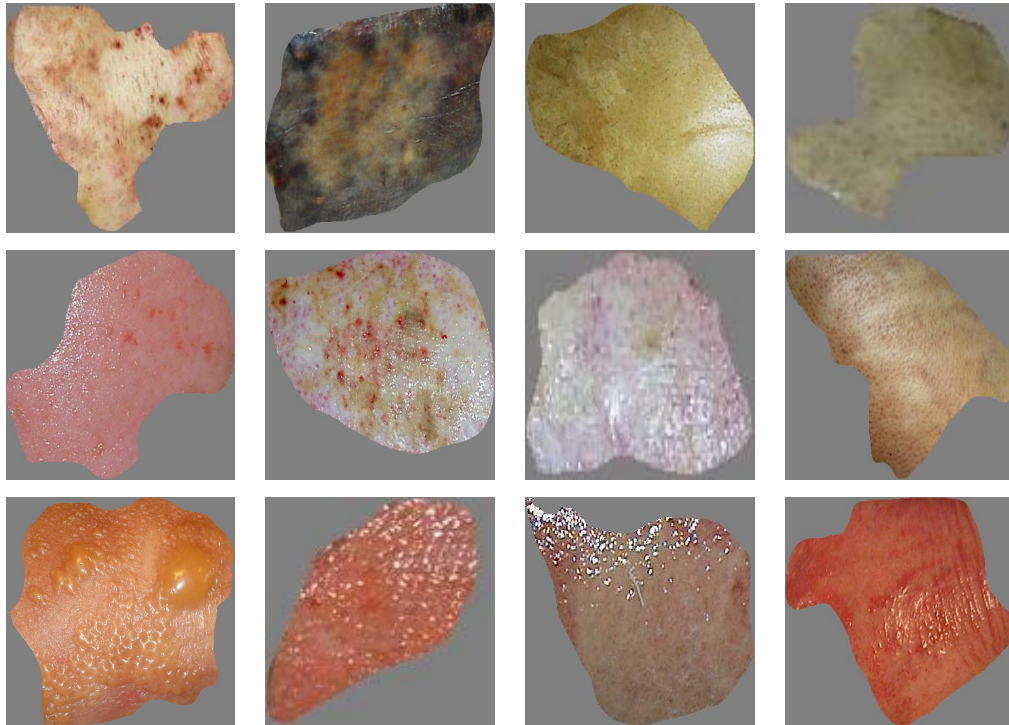
Figure 3.2: BIP_US database sample images. The rows from top to bottom display full-thickness, deep dermal, and superficial dermal samples, respectively.

since the goal was to compare our method with the approach used in [40], we followed the same data splitting strategy and put 74 images into the test set and the rest of the data samples into the train and validation sets.

Data Augmentation:

The training set was augmented using transformations such as rotating, flipping, cropping, and mirroring and 16 images were generated from each image after augmentation. Therefore, for the 3-class classification experiment, the number of training samples in the classes deep dermal, full-thickness, and superficial dermal increased to 416, 224, and 576 after augmentation. For the binary classification case, we ended up with 128 images in the non-grafted class and 144 images in the grafted category.

3.3.2 Training the DCNN using transfer learning

Due to the limited number of the images in the dataset and the complexity level of the classification problem, the AlexNet architecture was used in this research. AlexNet is a deep convolutional neural network proposed in 2012 as the winner of ILSVRC which outperformed all traditional machine learning methods [1, 49]. AlexNet consists of 8 layers, including 3 convolution and 2 fully connected layers, with 60 million parameters. This network is one of the most popular deep architectures in computer vision applications such as classification tasks [49]. We trained AlexNet using the transfer learning strategy. By using this method, an AlexNet that was pre-trained on ImageNet [75], was fine-tuned using the BIP_US dataset samples. ImageNet is a huge dataset that includes more than 14 million general images. All the experiments

in this study were implemented in version R2020a of MATLAB software. We used an Intel(R) Core (TM) i7-8565U CPU @1.80GHz 1.99 GHz and NVIDIA GEFORCE MX 150 GPU with 2GB of memory to run the experiments.

3.4   Results and discussion

Machine learning and deep learning have been used widely in the literature to solve healthcare problems including wound care challenges. Different DL-based methods were proposed for wound image analysis recently. However, the number of studies that cover the burn wound challenges including wound classification is limited. Also, in most cases the researchers discussed only the binary classification problem of burn wound images. In this article, we proposed an end-to-end DCNN-based strategy to classify the burn wound images into two and three categories. For the binary classification problem, we implemented the same experiment that was described in [40] for planned comparison. In this problem, the goal was to classify an input image into one of the two classes: the first class includes the burn wound images that required a grafting, and the second class included the non-grafted wound samples. Figure 3.3 shows the classification process in this experiment. The obtained results for this experiment have been summarized in Table 3.2 and Figure 3.4. We selected 0.0001 as the learning rate and trained the network for 10 epochs with the mini-batch size of 64. As we observe in Table 3.2, the test accuracy is 90.5%. Also, the recall, precision, F1-score, and AUC values are 87.9%, 90.6%, 0.8922, and 0.913, respectively.

For the 3-class classification experiment, the intent was to classify the input image
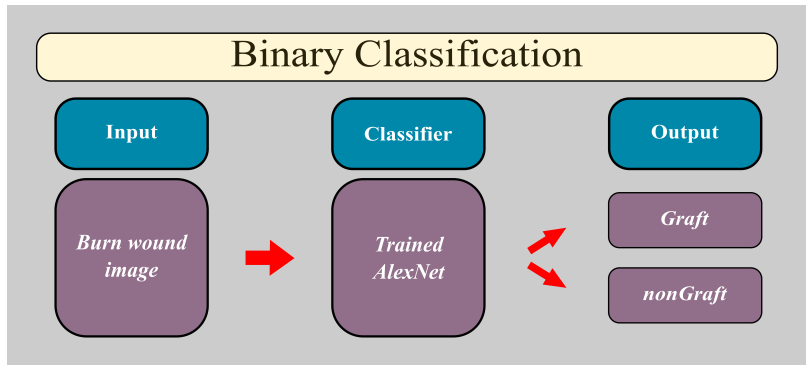
Figure 3.3: Binary classification process.

Table 3.2: Binary classification results

Classification	Accuracy (%)	Precision (%)	Recall (%)	F1-score	AUC
Graft/nonGraft	90.5	90.6	87.9	0.8922	0.913
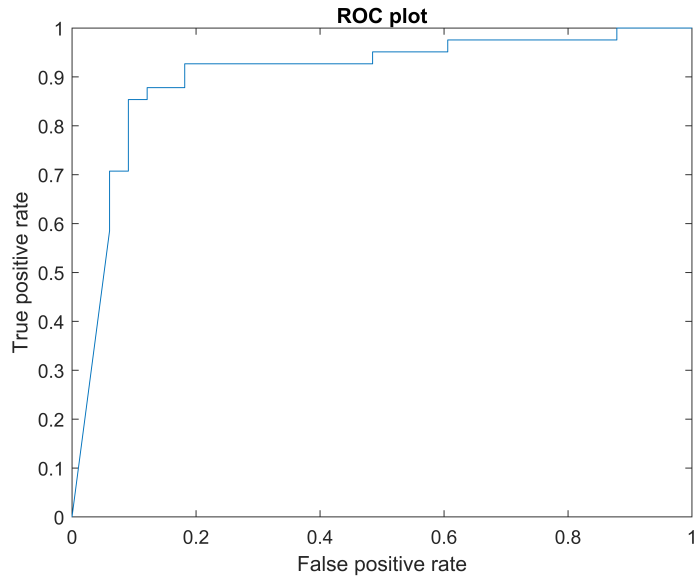
Figure 3.4: ROC plot for binary classification.

into one of the three classes: full-thickness, deep dermal, and superficial dermal. Figure 3.5 displays the classification process for this experiment. We used 1e-6 as

the learning rate value and trained the network for 5 epochs with the mini-batch size of 10. From the confusion matrix displayed in Figure 3.6, we observe that the test accuracy is 77.8%.

Regarding the results discussed above, we find that by increasing the number of classes from two to three, the classification accuracy decreased considerably. The justification is that by increasing the number of classes the number of network parameters will grow. It means that the network needs more data for training the new parameters, otherwise the classifier's performance will drop. Based on the reported results in [40] on binary classification, the authors obtained an accuracy value of 82.43% along with the precision, recall, and F1-score amount of 0.82, 0.88, and 0.85, respectively. As a conclusion, our designed classifier improved the binary classification accuracy by more than 8%. By considering all the reported metrics, we claim that our classifier generated better performance for the binary classification problem. For the 3-class classification case, the confusion matrix shows that the deep dermal is the easiest wound class to be classified by the network. The other two classes show the same difficulty level of classification. Currently no existing work has been seen on 3-class classification of burn wound images for comparison with the proposed work.

3.5   Conclusion

Wound image classification is one of the most important stages during the treatment process and a precise classifier can help the clinicians to have more efficient diagno-
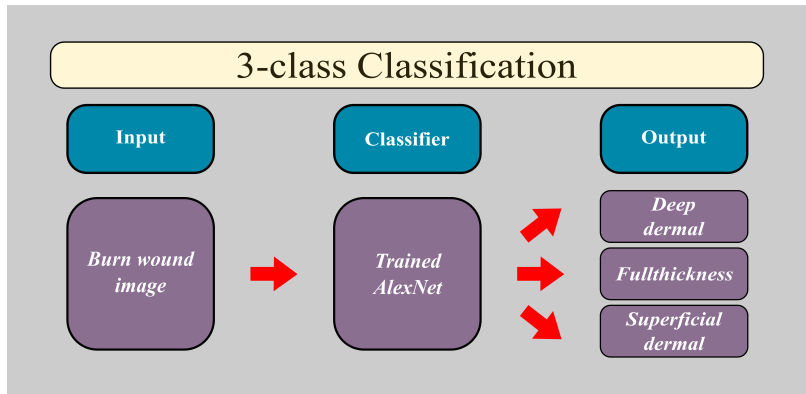
Figure 3.5: 3-class classification process.

Figure 3.6: Confusion matrix for 3-class classification problem.

sis. Many machine learning and deep learning-based methods have been presented in recent years to design high-performance burn wound image classifiers. However, most of the previous studies discussed only the binary classification problem. To fill this gap, in this research we presented a deep learning-based method for end-to-end classification of burn wound images into two and three classes. We used a pre-trained AlexNet and fine-tuned it using a burn wound image dataset, BIP_US. The results showed that our proposed approach can classify the burn wound images into two categories, grafted or non-grafted, satisfyingly. Also, our results display a considerable improvement over similar works in the literature. Moreover, despite having a very limited number of samples in the dataset, the proposed method provided a decent performance for classification of the burn wound samples into three categories: deep dermal, full-thickness, and superficial dermal. Both experiments in this study were limited by a small number of images and we expect that by using a larger dataset the results would be significantly improved. Our findings demonstrate that deep convolutional neural networks can be used successfully for burn wound image classification tasks or other similar clinical applications to improve the prognosis and treatment procedure.

Chapter 4

Wound Image Classification Using an Ensemble Classifier

4.1 Related works

In this section, we review the previous studies in wound image classification focusing on chronic wounds. We have organized the articles under two main subcategories: the papers which used a feature extraction method followed by an SVM, and the articles in which an end-to-end DCNN-based classification approach was proposed. The complete organization chart for the reviewed studies can be found in Figure 4.1.

4.1.1 Feature generation + SVM

Traditional ML algorithm + SVM

Yadav et al. proposed a method for binary classification of burn wound images using machine learning tools [40]. The authors classified the images into the two categories, grafting and non-grafted wounds, using a classic color-based feature extraction approach followed by an SVM. The dataset included 94 images with different burn depths including full-thickness, deep dermal, and superficial dermal. A classification accuracy of 82.43% was reported. Testing the proposed method on a very small

dataset is the weakness of this research.

Goyal et al. suggested detection and localization methods for Diabetic Foot Ulcer (DFU) on mobile devices [41]. For the classification part, they tried both conventional and DCNN-based methods. A dataset with 1775 DFU images was used and the ground truth generated by creating bounding boxes around the ROI using an annotation software. For the traditional machine learning techniques, patches were extracted from normal skin and abnormal areas and the number of samples increased by utilizing data augmentation methods. Different traditional feature extraction algorithms were applied, and the best three methods were selected. Then they extracted 209 features from each patch which were used for training a Quadratic SVM classifier. Finally, for a new image, sliding window technique was used for classifying each patch of the image as normal or abnormal by utilizing the trained SVM.
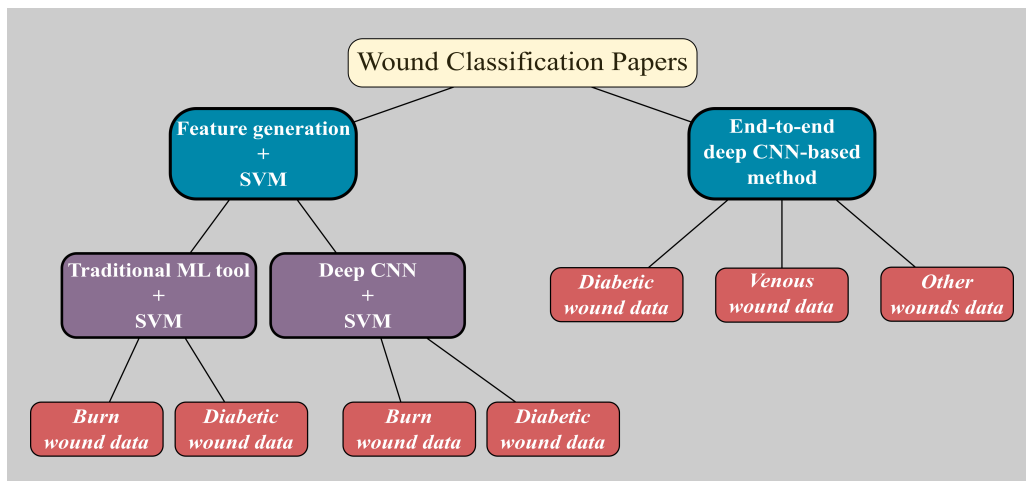
Figure 4.1: Organization chart for the wound classification papers.

Deep CNN + SVM

In another study, Abubakar et al. proposed a machine learning-based approach to distinguish between burn wounds and pressure ulcers [72]. Pre-trained deep architectures like VGG-face, ResNet101, and ResNet152 were utilized for feature extraction. The features were fed into an SVM for the classification task. The dataset included 29 pressure and 31 burn images which were augmented using cropping, rotation, and flipping transformations. After augmentation, they performed binary and 3-class classification experiments. In binary classification experiment, the images were classified into burn or pressure categories and in 3-class classification problem the goal was to classify the images into the labels burn, pressure, or healthy skin. ResNet152 architecture generated the best results for both classification problems with an accuracy value of 99.9%.

Goyal et al. predicted the presence of infection or ischemia in DFUs using a deep learning-based classification method [76]. A new dataset with 1459 DFU images was introduced and the samples were augmented using Faster-RCNN and Inception-ResNetV2 networks. Binary classification experiments were performed to classify the samples into infection or non-infection, and ischemia or non-ischemia classes. In more detail, some color-based descriptors were extracted from each patch before classification. ResNet50, InceptionV3, and InceptionResNetV2 architectures were used in this study. Besides, the authors used an ensemble CNN approach for combining the outputs of the three deep networks and fed it into an SVM for classification. They used MATLAB and TensorFlow frameworks. In both binary classification problems, the deep learning-based methods showed a better performance than the traditional

classifiers. The authors reported the accuracy values of 90% for ischaemia and 73% for infection experiments.

4.1.2 End-to-end deep CNN-based methods

In [41], for the deep learning-based classification methods, two-tier transfer learning approach was utilized for training the deep architectures including MobileNet, InceptionV2, ResNet101, and InceptionResNetV2. This method uses both partial and full transfer learning which means transferring only the lower level features or the whole features from a pre-trained model to the new model. Tensorflow was used as the framework. Object localization algorithms like R-FCN and Faster R-CNN were utilized followed by the trained deep architectures for tasks like classification. The combination of Faster R-CNN and InceptionV2 reported as the best model.

In another research, Goyal et al. used convolutional neural networks to classify diabetic foot ulcers [42]. A DFU image dataset with 397 images was presented. Data augmentation techniques were utilized to increase the number of samples. They proposed DFUNet, a deep neural network, for patch-wise classification of the foot ulcers into either normal or abnormal classes. DFUNet utilized the idea of concatenating the outputs of three parallel convolutional layers which used different filter sizes. The authors claimed that using this idea, multiple-level features were extracted from the input which resulted in having a network with higher discriminative strength. An accuracy value of 92.5% was reported for the proposed method. The main issue about this research is that the authors proposed a patch classifier which is not very helpful in medical image classification tasks. Indeed, it makes more sense to the

33

clinicians and is more useful, to work with a whole-image classifier instead of a patch classification model.

Nilsson et al. proposed a CNN-based method for venous ulcer image classification [43]. The utilized dataset included 300 samples and a VGG-19 network was used to classify the images into venous or non-venous categories. The methodology included pre-training of the VGG-19 network using another dataset of Dermoscopic images and then, fine-tuning the network utilizing their related dataset. The values obtained for accuracy, precision, and recall reported as 85%, 82%, and 75%, respectively. Caffe, TensorFlow, and keras were used as the frameworks.

Alaskar et al. applied deep convolutional neural networks for intestinal ulcer detection in wireless capsule endoscopy images [44]. AlexNet and GoogleNet architectures were utilized to classify the input images into ulcer (abnormal) or non-ulcer (normal) categories. The dataset consisted of 1875 images obtained from wireless capsule endoscopy video frames and the experiments implemented in MATLAB environment. The authors reported classification accuracy of 100% for both networks.

In another research, Shenoy et al. proposed a method for binary classification of wound images using deep CNNs [36]. A dataset with 1335 wound images collected via smartphones as well as the internet, was used in this study. After pre-processing and augmentation, nine different labels were created and for each label two positive and negative subcategories were considered. The authors created a modified form of VGG16 network, WoundNet, and three different versions of WoundNet were pre-trained on the ImageNet dataset. Besides, another network named Deepwound, which was an ensemble model was designed for averaging of the outcomes from

34

the three individual WoundNet architectures. The algorithms were implemented in Keras. Also, an application was created for mobile phones to facilitate patient to physician consultation and wound healing evaluation.

Alzubaidi et al. presented a DCNN for binary classification of diabetic foot ulcers [77]. A new dataset consisting of 754 smartphone-captured foot images was introduced in this study. The goal was to classify the samples into normal or abnormal (DFU) skin categories. Normal and abnormal patches were extracted from the images and number of samples increased using data augmentation techniques. The proposed network, DFU_QUTNet, is a deep architecture with 58 layers including 17 convolutional layers. In comparison with the common DCNNs, the width of the proposed model has been increased without adding computational complexities. Then the network would be able to extract more information from the input which results in higher classification accuracy. In one experiment, DFU_QUTNet was applied for an end-to-end classification task and in another one, it was utilized as a feature extractor along with SVM and KNN classifiers. The maximum reported F1-Score was 94.5% obtained from combining DFU_QUTNet and SVM. Although designing a high-performance patch classifier can be a good achievement, but in clinical environments it would be more useful to have a whole image classification system and it is the weakness of this research.

Table 4.1 summarizes the reviewed studies. Only a few papers were identified that discuss wound analysis from the wound type classification point of view. Also, most of the publications on wound type classification, discuss only the binary classification problems such as classifying the samples into normal and abnormal cases. Having

35

Table 4.1: Summary of Wound Image Classification Works.

Ref	Classification	Feature(s)	Methods	Dataset	Limitation(s)
[42]	Binary (DFU/normal skin).	N/A	A novel CNN architecture named DFU-Net.	DFU dataset with 397 images (292 abnormal, 105 normal cases).	For small DFUs and DFUs having similar color like surrounding skins is hard to classify by this network. This also goes for normal skins with wrinkle and high red tones.
[40]	Binary (graft/non-graft burn wound)	color, texture, and depth.	Support Vector Machine (SVM).	Burns BIP_US Database.	Very small evaluation set (74 images) was used.
[43]	Binary (venous/non-venous)	N/A	A pre-trained VGG-19 network.	A dataset with 300 images with specialist annotation.	Classification accuracy depends on camera distance from the ulcer surface.
[44]	Binary (ulcer/non-ulcer endoscopy image)	N/A	AlexNet and GoogleNet.	1875 images obtained from wireless capsule endoscopy video frames.	An unbalanced test set (3:1 ratio) has been used.
[36]	Binary (considered pos/neg cases for different labels like wound, infection, granulation, etc.	N/A	WoundNet (modified version of VGG-16) and Deepwound (an ensemble model).	1335 wound images collected via smartphones and internet.	As accuracy varies from 72% to 97% for different binary classes, for some specific classes (like: drainage (72%)) this model does not work well.
[72]	Binary (burn/pressure), 3-class (burn/pressure/healthy skin)	Feature extracted from VGG-face, ResNet101, and ResNet152.	Support Vector Machine (SVM).	29 pressure and 31 burn images.	Very small dataset used.
[77]	Binary (normal/abnormal skin (diabetic ulcer)).	N/A	A novel deep CNN (DFU_QUTNet).	754 foot images.	Only use precision, recall and f1-score as evaluation matrices, which may not reflect all the evaluations clearly.
[76]	Binary (infection/non-infection, ischemia/non-ischemia)	Bottleneck features extracted from Inception-V3, ResNet50, and Inception-ResNetV2.	Support Vector Machine (SVM).	1459 DFU images.	Depending on lighting conditions (shadow), marks, and skin tone their model can show poor performance.
[41]	Binary (normal/abnormal skin (DFU))	209 features extracted using LBP, HOG, and color descriptors.	Quadratic SVM, InceptionV2, MobileNet, ResNet101, Inception-ResNetV2	1775 DFU images collected from a hospital within five years.	No evaluation of their classification task has been given.

difficulties to access a reliable dataset can be mentioned as a reason for this issue. Providing data to fill this gap in the literature was one of the motivations for our research. Moreover, many papers discussed only the patch-wise or ROI classification instead of the image-wise wound classification. In the rest of this section, we propose an ensemble classifier for image-wise multi-class wound type classification using deep convolutional neural networks.

4.2   AZH dataset

In this research, we used a new wound image dataset collected over a two-year clinical period at the AZH Wound and Vascular Center in Milwaukee, Wisconsin. The dataset includes 400 wound images in jpg format and various sizes in the range 240×320 to 525×700 and bit depth of 24 from four different wound types: venous, diabetic, pressure, and surgical (100 images per class). The images were captured using an iPad Pro (software version 13.4.1) and a Canon SX 620 HS digital camera and were labeled by a wound specialist from the AZH Wound and Vascular Center. Except than a few cases, each wound image is from a separate patient. For those exceptional cases we took more than one image of the same patient but from different parts of the body or in different healing stages. The dataset can be accessed on GitHub with this link [https://github.com/uwm-bigdata/wound_classification]. Figure 4.2 shows some sample images from different classes of the dataset.

Figure 4.2: Sample images from the AZH Wound and Vascular Center database. The rows from top to bottom display diabetic, venous, pressure, and surgical samples, respectively.

4.3 Method

This subsection describes the method we used in this research. Figure 4.3 displays how the proposed ensemble classifier works. As it can be found from this figure, the proposed classifier utilizes two different whole image classifiers (A and B) at the same time which work using two different classification strategies: patch-wise and image-wise classification. In this figure, C_1 to C_n show Class 1 to Class n in which n is the number of classes. For the binary and three-class classification problems, n would be 2 and 3, respectively. The input image is fed into both classifiers A and B. Then the output of these classifiers are combined and fed into a MLP classifier to predict the class of the whole input wound image. It is important to mention that the idea behind combining the Classifier A and B is to consider both patch level and whole image level information for classification. In the next two subsections, we have provided the details for different components of the proposed algorithm.

4.3.1 Patch-wise classification

Pre-processing-ROI Extraction

We selected 100 images for each wound type and manually extracted 100 unique Regions of Interest (ROI) out of them per class, representing each of the six categories: diabetic, venous, pressure, surgical, background, and normal skin. The ROIs are rectangular and have different sizes. Figure 4.4 displays some of the extracted ROIs from different classes of the dataset.
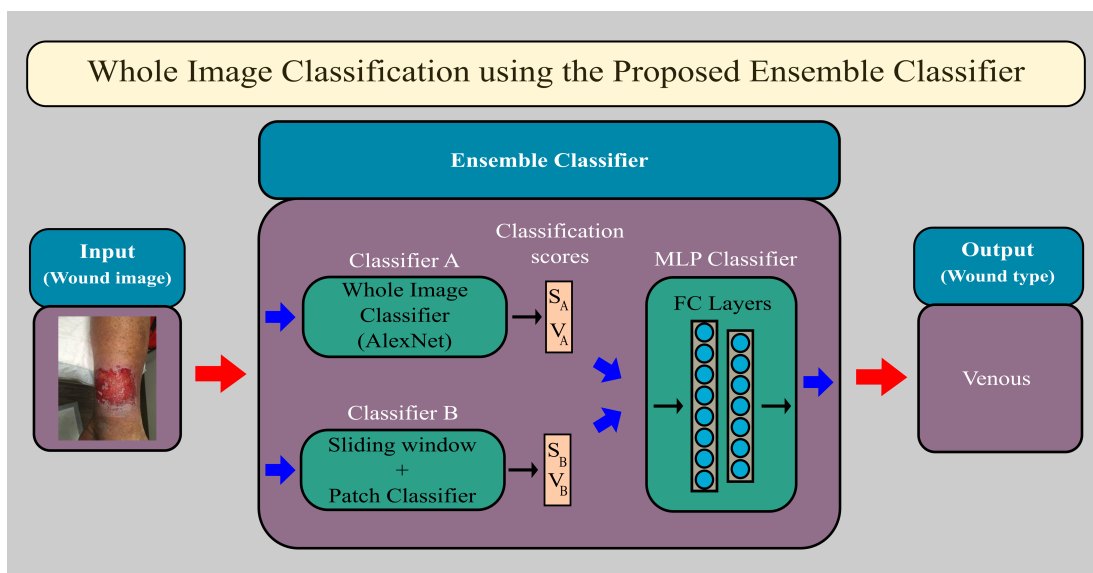
Figure 4.3: Whole image classification process using our proposed ensemble classifier. The classifier accepts the wound image as the input and predicts the wound type as the output.

Pre-processing-Data splitting

After extracting the ROIs, we put them randomly into training (70% samples), testing (15% samples), and validation (15% samples) sets. None of the training images, was used in the test set and vice versa.

Pre-processing-Patch generation

In this step, 17 patches were generated from each ROI. The patches were extracted in a way that covers between 75% to 85% of the original ROI. These are rectangular area extracted from different regions of ROI using cropping. The patches were extracted in a way that cover different regions of the ROI including center region, up-left region, down-left region, up-right region, down-right region, and so on. By try and error, it

40
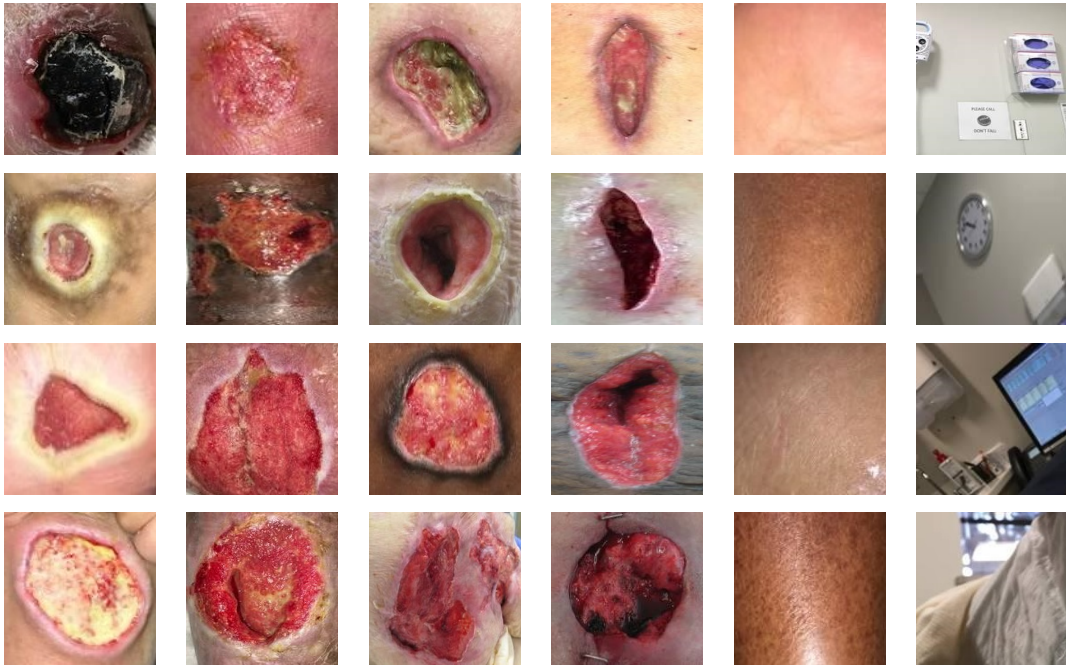
Figure 4.4: Sample ROIs. The columns from left to right display diabetic, venous, pressure, surgical, normal skin, and background ROIs, respectively.

was found that by generating patches from 17 different regions of each ROI we can have better training results for the classifier B. After this step, for each class we had 1190, 255, and 255 patches in the train, validation, and test set, respectively.

Pre-processing-Data Augmentation

Augmentation of the training set samples was performed by generating 16 samples from each one using image transformation methods like rotating, flipping, cropping, and mirroring. Following augmentation, there were 19040 training samples in each class.
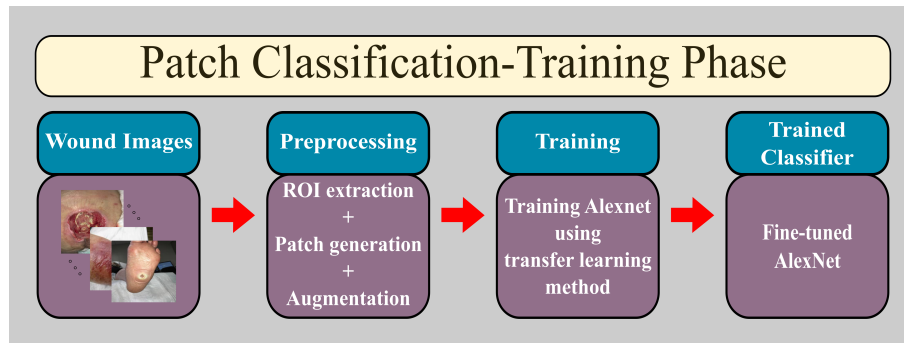
Figure 4.5: Training process of the patch classifier.

Training

Following the pre-processing step, we trained a deep convolutional neural network using the training samples. After studying various networks, since our training set is small and the more modern CNNs need extensive number of samples and are computationally expensive, we selected AlexNet as the classifier. Due to its simplicity and effectiveness, AlexNet is still one of the most common deep networks used by

researchers [78]. As it's mentioned in the literature [79], AlexNet's training time is considerably lower than other complex architectures. Moreover, using more complex networks needs more memory [79] which is an important factor in choosing the proper network. AlexNet is a deep CNN architecture proposed in 2012 and was the winner of ILSVRC in comparison to other traditional machine learning methods [1, 49]. This network contains eight layers including five convolutional and three fully connected layers and has 60 million parameters. Also, max-pooling layers have been used after the first, second, and fifth convolutional layer. A modified version of the original AlexNet described above was utilized in this research to suit for our classification problems. We changed the last fully connected layer in a way that its output size, matches the number of classes in our data. We used this modified version of AlexNet for all of our experiments and classifiers in this study. In addition, we utilized the transfer learning technique to increase the training accuracy while reducing the training time. It means that the AlexNet architecture was pre-trained on a massive dataset of general images called ImageNet and fine-tuned using our wound image patches. Figure 4.5 shows the described steps for training the patch classifier.

4.3.2 Image-wise classification using an ensemble classifier

Several DCNN-based ensemble classification methods were proposed for medical or non-medical image classification tasks in the literature [80, 81, 82]. Various strategies were used by the researchers to construct the ensemble classifier such as voting, concatenating, averaging, etc. [83, 84, 85, 86]. In all of these studies, the final conclusion was that the ensemble model outperformed the individual classifiers in performance.

To this end, we designed an ensemble classifier in which the trained patch classifier described in Section 4.3.1 is used as a building block. In fact, the classification scores acquired from two classifiers (patch-wise and image-wise) are fed into a Multilayer Perceptron (MLP) classifier to obtain a better classification performance. Our hypothesis is that the proposed ensemble classifier will outperform each of the individual classifiers in terms of classification accuracy. Different components of the proposed classifier will be explained individually below.

Classifier A - Whole image classifier

This classifier is a pre-trained AlexNet architecture that we fine-tuned using our own dataset. Figure 4.6 displays the training phase for this classifier.
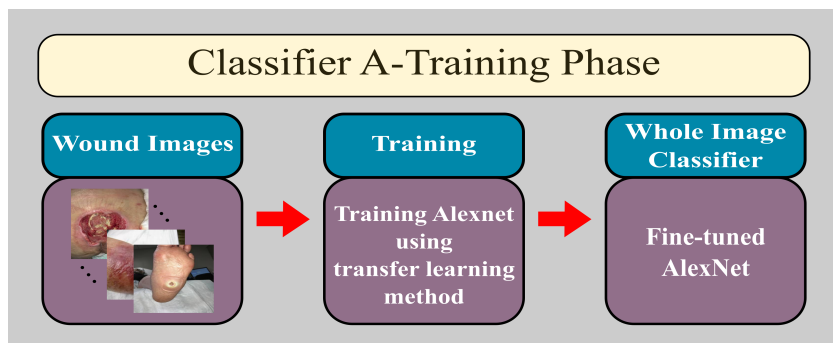
Figure 4.6: Training process of Classifier A.

Classifier B - Sliding window + Patch classifier

This classifier applies the sliding window technique on the input wound image to extract 9 patches of equal size along with patch classification step to predict their wound type. The wound type for the whole image will then be predicted by majority

voting on the predicted label of the patches detected as wound. Figure 4.7 describes the entire process for this classifier.
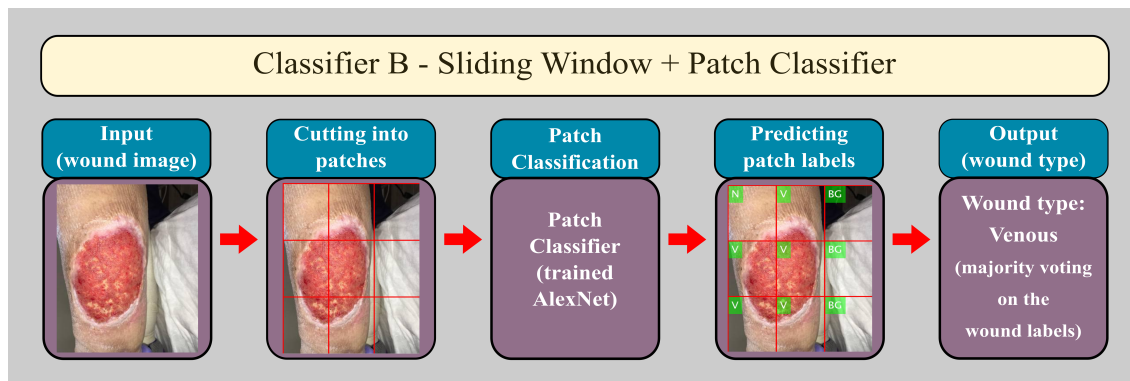
Figure 4.7: Classifier B. The first step is extracting equal size patches out of the input image using the sliding window technique. Then the patch classifier is used for detecting the patch labels. The final step is majority voting for predicting the whole image label.

Classification scores

For every input image, each of the classifiers A and B generate classification scores for all classes. For example, for the surgical vs venous classification problem, they generate S and V scores which stand for classification scores of the surgical and venous labels, respectively (see Table 4.2 for abbreviations). For the classifier B, these scores are calculated by averaging over S and V scores of all the patches detected as wounds by the patch classifier. In the end, for the binary classification case, we create a three-element feature vector including S_A, S_B, and V_B in which the subscripts A and B show the related classifier. It is important to note that we did not include V_A in the feature vector, because of the correlation between S_A and V_A. Finally, we feed

the feature vector into the MLP classifier for the final classification task as described below.

MLP Classifier

The MLP classifier is a four-layer MLP with two hidden layers that have 8 and 7 neurons, respectively. The number of nodes in the input and output layers are determined based on the type of the classification problem. The output of the MLP classifier is the wound type of the input image.

4.3.3   Performance metrics

In this research we used accuracy, precision, recall, and F1-score metrics to investigate the performance of the classifiers. Equations 2.1 to 2.4 show the related formulae for these evaluation metrics. In the binary classification problem, we used Area Under the ROC Curve (AUROC or AUC) metric as well. TP, TN, FP and FN in the mentioned equations stand for True Positive, True Negative, False Positive, and False Negative measures, respectively. More details about these equations and the related concepts can be found in [68].

4.4   Results

This subsection presents the results obtained from the patch-wise and image-wise classification experiments. The classifiers were implemented in MATLAB R2019b and R2020a using an NVIDIA GEFORCE RTX 2080 Ti GPU with 11GB of memory.

Abbreviation	Description
BG	Background
N	Normal skin
V	Venous
D	Diabetic
P	Pressure
S	Surgical

Table 4.2: Class label abbreviations

In the diagrams and tables presented in this section, the following abbreviations D, V, P, S, BG, and N represent the classes diabetic (D), venous (V), pressure (P), surgical (S), background (BG), and normal skin (N) (Table 4.2). Several experiments were conducted to find the optimum training parameters of the deep networks. The optimum epoch number obtained was 20 and we used a learning rate value of 10e-6. Also, Adam was utilized as the optimization algorithm [87]. Further details for each experiment are provided below.

4.4.1   Patch classification

To evaluate the patch classifier's performance for patch-wise classification, we used 255 test patches per class. For 4-class classification experiments, the goal was to classify the wound patches into one of the four classes: BG, N, and two wound labels. In the 5-class classification problem, we had three wound labels as well as the BG and N classes. The last group of the patch-wise classification experiments is related to the 6-class classification case in which the wound patches are classified into one of the six classes diabetic, venous, pressure, surgical, BG, and N. Table 4.3 shows the test accuracy values for all the experiments mentioned above. Figures 4.8 to 4.15

display some sample confusion matrices for patch-wise classification experiments. It should be noted that we performed and compared all the experiments with and without data augmentation. As data augmentation always resulted in better results, we only show our experiments with data augmentation.

Num of Classes	Classes	Test accuracy (%)
4-class	BGNVD	89.41
	BGNVP	86.57
	BGNVS	**92.20**
	BGNDP	80.29
	BGNDS	90.98
	BGNPS	84.12
5-class	BGNDVP	79.76
	BGNDVS	**84.94**
	BGNDPS	81.49
	BGNVPS	83.53
6-class	BGNDVPS	**68.69**

Table 4.3: Patch-wise classification results.

4.4.2 Whole image classification

To assess our proposed ensemble classifier's efficiency, we performed two types of experiments: binary classification and 3-class classification. With the patch classification results presented in subsection 4.4.1, surgical vs venous and surgical vs venous vs diabetic classifiers were selected which showed the best binary and 3-class classification outcomes. For the rest of the chapter, we name the whole image classifier (trained AlexNet on the whole wound images) as Classifier A, and the other classifier which uses patch classification strategy (sliding window + trained AlexNet on the wound patches) as Classifier B. We also obtained 138 extra wound images from three

Figure 4.8: Confusion matrix for 4-class classification (BGNVD) experiment.

classes surgical (28 samples), diabetic (54 samples), and venous (56 samples) from the AZH wound and vascular center to be used as the test images. Table 4.4 shows the binary classification results obtained from applying the Classifier A, Classifier B, and the proposed ensemble classifier on the test set which included 84 wound images. Also, the 3-class classification results for the three classifiers have been provided in Table 4.5. The superiority of our proposed ensemble classifier over the other two classifiers can be detected from these results for both binary and 3-class classification problems.

Figure 4.9: Confusion matrix for 4-class classification (BGNDS) experiment.

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
A	83.3	76.9	71.4	74.04
B	82.1	71	78.6	74.60
Our ensemble classifier	**96.4**	**93.1**	**96.4**	**94.72**

Table 4.4: Whole image binary classification (surgical vs venous) results obtained from applying the classifiers on the test set images.

4.5 Discussion

In the patch classification, by looking at Table 4.3 and Figures 4.13 and 4.14 as we expected, by increasing the number of classes from four to five and six, the classifica-

Figure 4.10: Confusion matrix for 4-class classification (BGNPS) experiment.

Classifier	Class	Precision (%)	Recall (%)	F1-score (%)	ACC (%)
A	D	88	81.5	84.62	83.3
	S	67.9	67.9	67.9	
	V	86.7	92.9	89.69	
B	D	70.9	72.2	71.54	67.4
	S	42.1	28.6	34.06	
	V	71.9	82.1	76.66	
Our ensemble classifier	D	86.2	92.6	89.28	89.1
	S	81.5	78.6	80.02	
	V	96.2	91.1	93.58	

Table 4.5: Whole image 3-class classification results obtained from applying the classifiers on the test set images.

Confusion Matrix

	BG	D	N	P	S	
BG	**221** 17.3%	**0** 0.0%	**2** 0.2%	**0** 0.0%	**5** 0.4%	96.9% 3.1%
D	**0** 0.0%	**205** 16.1%	**0** 0.0%	**43** 3.4%	**32** 2.5%	73.2% 26.8%
N	**34** 2.7%	**0** 0.0%	**249** 19.5%	**0** 0.0%	**0** 0.0%	88.0% 12.0%
P	**0** 0.0%	**31** 2.4%	**0** 0.0%	**181** 14.2%	**35** 2.7%	73.3% 26.7%
S	**0** 0.0%	**19** 1.5%	**4** 0.3%	**31** 2.4%	**183** 14.4%	77.2% 22.8%
	86.7% 13.3%	80.4% 19.6%	97.6% 2.4%	71.0% 29.0%	71.8% 28.2%	**81.5% 18.5%**

Figure 4.11: Confusion matrix for 5-class classification (BGNDPS) experiment.

tion accuracy decreased. The justification for this phenomenon is that increasing the number of classes accordingly increases the number of network parameters, which would make it more challenging for the deep architecture to train all the parameters to the same standard as before, using the same number of training samples. Another interesting observation is that in the 4-class and 5-class classification experiments, the lowest classification accuracy is related to the diabetic and pressure wounds. It shows that these two wound types are very similar in appearance. The confusion matrices confirm this fact by showing that many diabetic wounds were classified into pressure class and vice versa. As another observation, we see that the surgical and venous wounds are distinguishable with an acceptable accuracy level. In addition, we see that in most of the experiments, the pressure wound is the most challenging

Figure 4.12: Confusion matrix for 5-class classification (BGNVPS) experiment.

wound type to classify. By looking closely at the confusion matrices, we find that the low recall value for this wound type often comes from misclassifying the wound into the venous or diabetic instead of the pressure class. Another observation is that the venous wounds typically show the highest recall values among all of the wound types. This phenomenon could be related to having samples from a wider variability and consequently better training of the classifier for this wound type. We expect that increasing the number of samples for dataset categories would improve the recall value for all the wound types. Recall is an important factor when we are investigating the performance of a classifier. In all of the patch classification experiments, the background and normal skin classes have the highest recall values. This is important because in our proposed ensemble classifier we need to have a patch classifier with
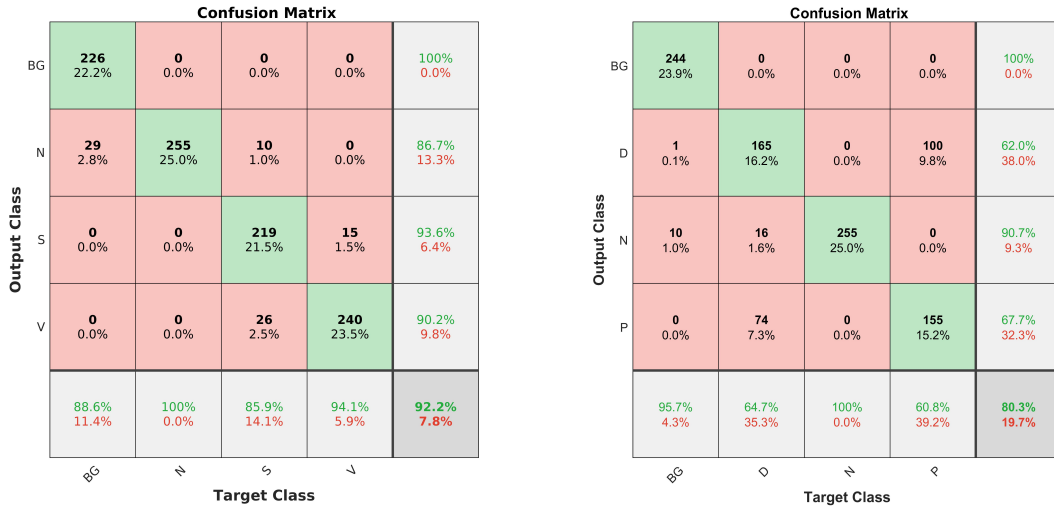
Figure 4.13: Confusion matrices of the best (left) and worst (right) case in 4-class classification experiments.
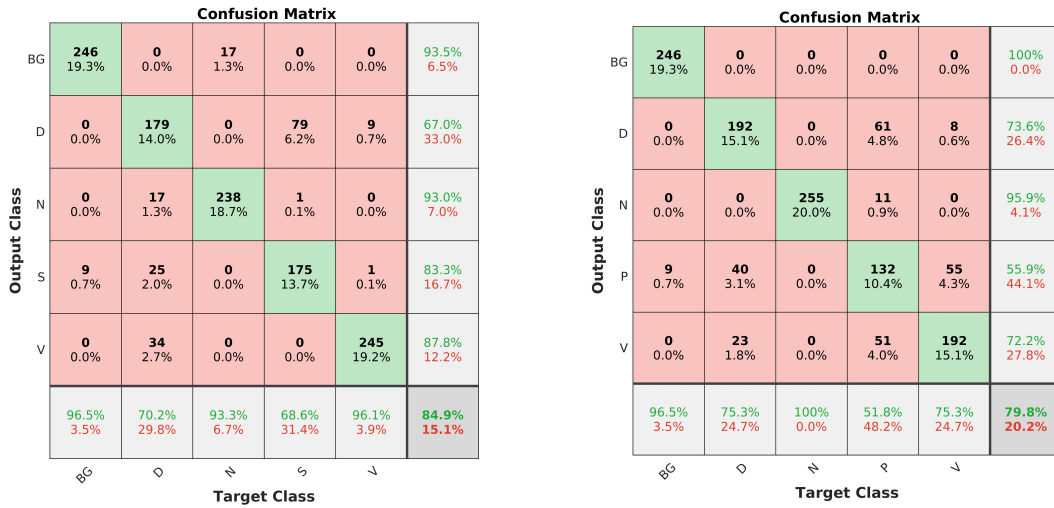
Figure 4.14: Confusion matrices of the best (left) and worst (right) case in 5-class classification experiments.

Confusion Matrix

	BG	D	N	P	S	V	
BG	**223** 14.6%	**0** 0.0%	**0** 0.0%	**0** 0.0%	**5** 0.3%	**0** 0.0%	97.8% 2.2%
D	**0** 0.0%	**167** 10.9%	**5** 0.3%	**81** 5.3%	**50** 3.3%	**17** 1.1%	52.2% 47.8%
N	**32** 2.1%	**0** 0.0%	**250** 16.3%	**0** 0.0%	**0** 0.0%	**1** 0.1%	88.3% 11.7%
P	**0** 0.0%	**13** 0.8%	**0** 0.0%	**121** 7.9%	**59** 3.9%	**42** 2.7%	51.5% 48.5%
S	**0** 0.0%	**32** 2.1%	**0** 0.0%	**1** 0.1%	**106** 6.9%	**11** 0.7%	70.7% 29.3%
V	**0** 0.0%	**43** 2.8%	**0** 0.0%	**52** 3.4%	**35** 2.3%	**184** 12.0%	58.6% 41.4%
	87.5% 12.5%	65.5% 34.5%	98.0% 2.0%	47.5% 52.5%	41.6% 58.4%	72.2% 27.8%	**68.7% 31.3%**

Figure 4.15: Confusion matrix for the 6-class classification experiment.

the ability to distinguish background and normal skin parts from the wound tissue with a high accuracy value.

About the image-wise classification experiments, by looking at the Tables 4.4 and 4.5 we see that for the binary case, both Classifier A and B generate almost similar results while our proposed ensemble classifier showed an accuracy value of 96.4% which is 13.1% higher than the Classifier A and 14.3% higher than the Classifier B. Also, for the 3-class classification case, the accuracy of the ensemble classifier is higher than the other two classifiers. This last observation is very interesting because the Classifier B displays a low classification performance specifically for the surgical wounds, but after combining with the Classifier A, improves its accuracy value by 5.8%.

4.5.1 Robustness experiments

To investigate the robustness of our proposed classifier, we used 5-fold cross-validation as a standard evaluation method. For these experiments, we added the test set (84 images for binary and 138 images for 3-class classification experiment) to our training set (200 images for binary and 300 images for 3-class classification case). Thereafter, for each of the classes we partitioned the images into five folds randomly. Then we selected the first fold (20% of the samples) as the test set and the rest as the training samples. We trained our classifier using the training samples and tested it on the test set. Next time, we selected the second fold as the test samples and trained the network on the remaining folds, and so on. Tables 4.6 to 4.10 display the classification accuracy, AUC, precision, recall, and F1-score values obtained for the binary classification problem. Figures 4.16 and 4.17 compares the classifiers in accuracy and AUC metrics. The ROC plots are presented in Figure 4.18. Also, 3-class classification results have been summarized in Table 4.11 to 4.14 and Figure 4.19. In all tables, R1 to R5 display the round number of the experiments.

Table 4.6: Whole image binary classification (S vs V) accuracy percentages obtained from 5-fold cross-validation.

Classifier	R 1	R 2	R 3	R 4	R 5
A	91.1	89.3	87.5	85.7	85.7
B	67.9	69.6	73.2	83.9	75
Our ensemble classifier	**94.6**	**94.6**	**96.4**	**92.9**	**92.9**

Regarding the reported results, we find that our proposed ensemble classifier beats both Classifiers A and B with displaying better classification performance. This find-

Table 4.7: Whole image binary classification (S vs V) AUC values obtained from 5-fold cross-validation.

Classifier	R 1	R 2	R 3	R 4	R 5
A	0.9497	0.9677	0.9548	0.9548	0.9303
B	0.7806	0.7716	0.7677	0.8439	0.7084
Our ensemble classifier	**0.9806**	**0.9845**	**0.9613**	**0.9561**	**0.9535**
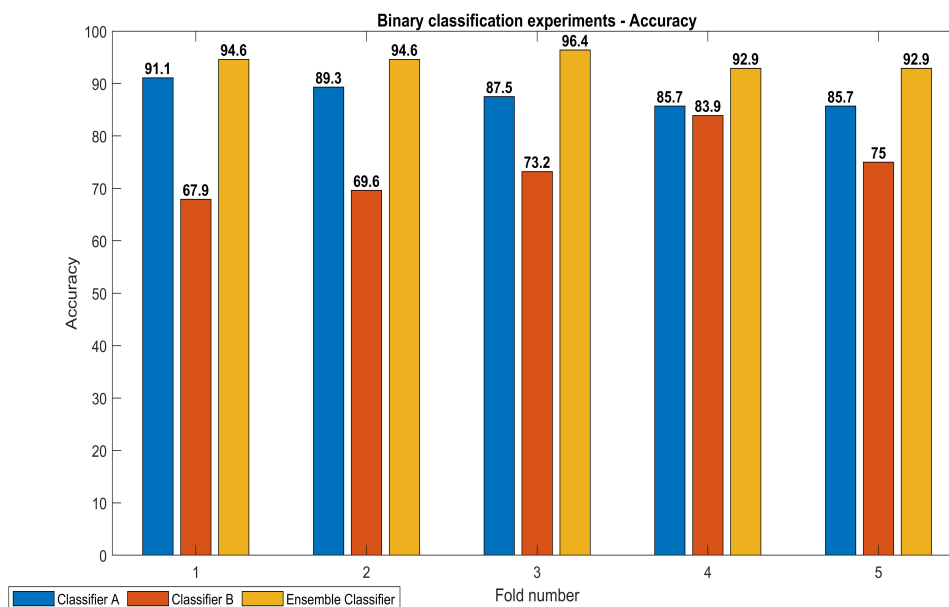
Figure 4.16: Accuracy values obtained from 5-fold cross-validation for the binary classification problem.

Table 4.8: Whole image binary classification (S vs V) Precision percentages obtained from 5-fold cross-validation.

Classifier	R 1	R 2	R 3	R 4	R 5
A	91.7	91.3	100	84	87
B	65.2	68.2	77.8	83.3	76.2
Our ensemble classifier	**95.8**	**89.3**	**100**	**86.2**	**92**
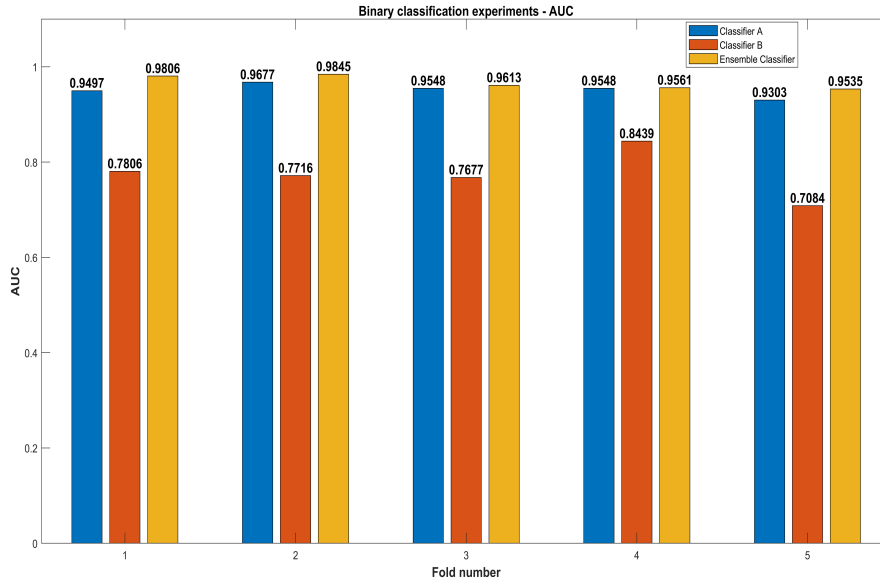
Figure 4.17: AUC values obtained from 5-fold cross-validation for the binary classification problem.

Table 4.9: Whole image binary classification (S vs V) Recall percentages obtained from 5-fold cross-validation.

Classifier	R 1	R 2	R 3	R 4	R 5
A	88	84	72	84	80
B	60	60	56	80	64
Our ensemble classifier	**92**	**100**	**92**	**100**	**92**

Table 4.10: Whole image binary classification (S vs V) F1-score percentages obtained from 5-fold cross-validation.

Classifier	R 1	R 2	R 3	R 4	R 5
A	89.81	87.49	83.72	84	83.35
B	62.49	63.83	65.12	81.61	69.56
Our ensemble classifier	**93.86**	**94.34**	**95.83**	**92.58**	**92**
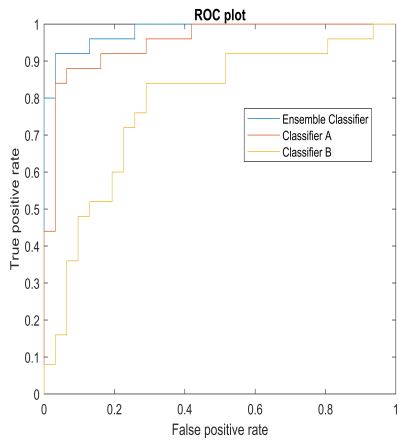
(a) Round 1
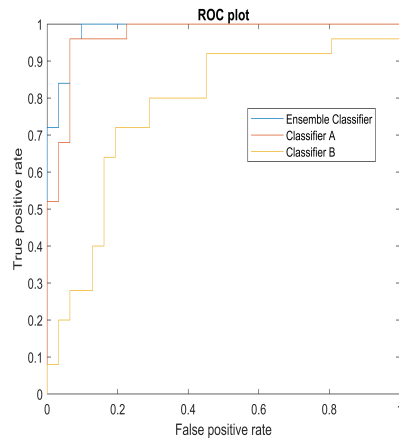
(b) Round 2
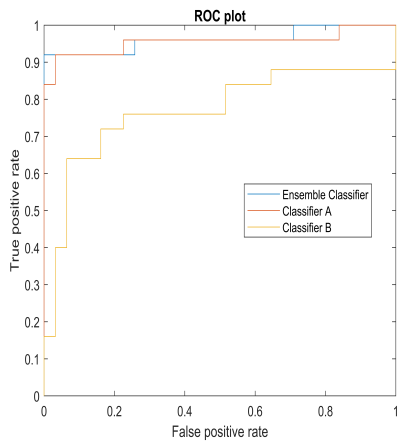
(c) Round 3
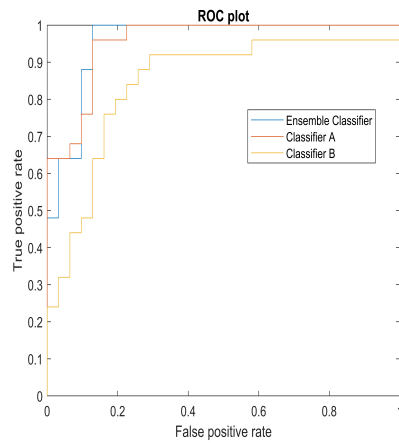
(d) Round 4
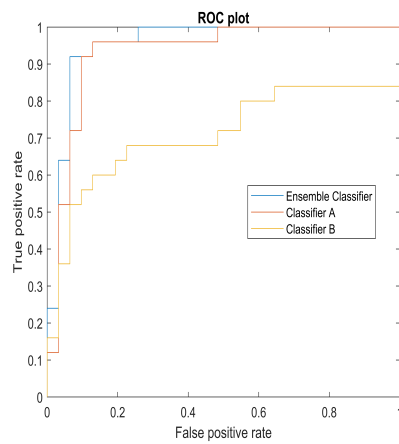
(e) Round 5

Figure 4.18: ROC plots obtained from 5-fold cross-validation experiments.

Table 4.11: Whole image 3-class classification (S vs V vs D) accuracy percentages obtained from 5-fold cross-validation.

Classifier	R 1	R 2	R 3	R 4	R 5
A	79.1	76.7	82.6	83.7	81.4
B	68.6	55.8	64	72.1	61.6
Our ensemble classifier	**84.9**	**81.4**	**88.4**	**91.9**	**91.9**
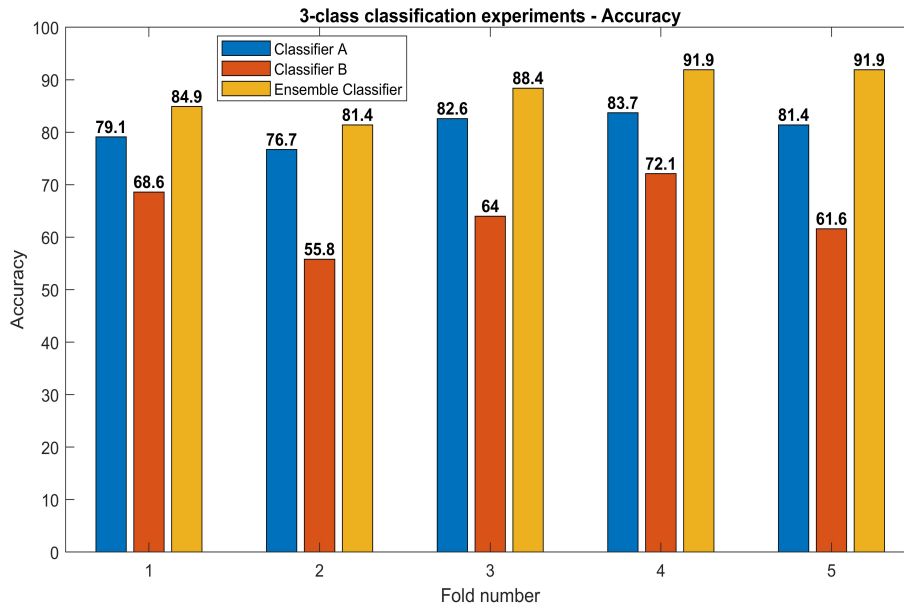
Figure 4.19: Accuracy values obtained from 5-fold cross-validation for the 3-class classification problem.

ing confirms our initial theory that combining two individual classifiers that one of them includes the patch-level information, and the other one contains the image-wise information will result in a stronger classifier with higher classification accuracy and better performance. Figure 4.20 presents some of the samples that were classified wrongly by Classifier A or B, but the proposed ensemble classifier assigned them to

Table 4.12: Whole image 3-class classification (S vs V vs D) precision percentages obtained from 5-fold cross-validation.

Classifier	Class	R 1	R 2	R 3	R 4	R 5
A	D	80	80	83.3	81.3	78.8
	S	73.9	76.2	84.2	94.7	80
	V	81.8	74.3	81.1	80	84.8
B	D	63.6	70.6	68	67.7	56.7
	S	65	41.4	44	66.7	63.2
	V	75.8	60	75	79.4	64.9
Our ensemble classifier	D	81.8	92.6	84.4	89.7	85.3
	S	85.7	70.4	90.5	100	91.7
	V	87.5	81.3	90.9	87.9	100

Table 4.13: Whole image 3-class classification (S vs V vs D) recall percentages obtained from 5-fold cross-validation.

Classifier	Class	R 1	R 2	R 3	R 4	R 5
A	D	80	80	83.3	86.7	86.7
	S	68	64	64	72	64
	V	87.1	83.9	96.8	90.3	90.3
B	D	70	40	56.7	70	56.7
	S	52	48	44	56	48
	V	80.6	77.4	87.1	87.1	77.4
Our ensemble classifier	D	90	83.3	90	86.7	96.7
	S	72	76	76	96	88
	V	90.3	83.9	96.8	93.5	90.3

the correct class. The interesting observation was that the major part of the samples miss-classified by Classifier A, were the images in which the wound consists a small proportion of the entire image. On the other hand, Classifier B missed the samples in which the wound occupies a large part of the image. These observations show that how the two weak classifiers cooperate to fix each other's shortcomings which resulted in producing a superior classifier. Indeed, having objects from different scales in the dataset has always been one of the challenges in deep learning-based

Table 4.14: Whole image 3-class classification (S vs V vs D) F1-score percentages obtained from 5-fold cross-validation.

Classifier	Class	R 1	R 2	R 3	R 4	R 5
A	D	80	80	83.3	83.91	82.56
	S	70.82	69.56	72.72	81.80	71.11
	V	84.36	78.80	88.25	84.83	87.46
B	D	66.64	51.06	61.83	68.83	56.7
	S	57.77	44.45	44	60.88	54.56
	V	78.12	67.59	80.59	83.07	70.60
Our ensemble classifier	D	85.70	87.70	87.11	88.17	90.64
	S	78.25	73.09	82.61	97.95	89.81
	V	88.87	82.57	93.75	90.61	94.90

tasks as discussed in several studies [88, 89, 90]. Specifically, in the field of wound care it is not guaranteed to take high quality images from an optimal view point and desired distance to the wound surface because of medical concerns like infection control as well as the patient's easement [42]. Our results show that the Classifier B can overcome partially the scale problem for those images which were taken from a further distance to the wound. It of course loses this power for the photos zoomed on the wound area. It should be mentioned that we conducted the experiments for 4-class whole image classification problem as well and obtained average and maximum accuracy percentages of 65.48 and 68.9, respectively. Regarding these results and the results presented in Figure 11, we didn't continue our experiments with this classification case as the final accuracy value would be low. We believe that having a limited number of data samples can be an important factor to get worst performance for four classes. Then we limited this study to only binary and three class classification and left the higher number of classes for a future study with a larger dataset.

Figure 4.20: Some of the miss-classified samples. The top row shows the samples miss-classified by Classifier B, and the bottom row displays the images which were wrongly classified by Classifier A.

4.5.2 Comparison with other models

For further investigation of our proposed model, we compared its performance with some of the models that people used in the literature. ResNet architecture is one of the common networks used in the wound image classification field in the recent years [71, 72, 76, 41]. We selected ResNet101 and ResNet50 as two different deeper and shallower versions of the ResNet family. Also, we compared our model with SqueezeNet architecture which generated the similar accuracy level of AlexNet on ImageNet dataset [91]. Table 4.15 shows the comparison results for both binary and 3-class classification experiments on our own dataset. For each classifier in the table, Avg and Max show the average and maximum accuracy values obtained from 5-fold cross validation experiments. Also, time column, displays the approximate training time for each model. For our model, this time is related to the Classifier B that takes longer time to be trained. As it can be found from these results, our proposed

63

ensemble classifier showed a better performance in both Avg and Max accuracy values.

4.5.3 Applying on Medetec dataset

As an additional evaluation, we applied our proposed ensemble classifier on Medetec images which is a famous public available wound image dataset [92]. For the binary classification experiment we classified the pressure wounds versus venous-arterial ulcers and in 3-class classification case, we added the diabetic category as well. The dataset includes 177 images in pressure class, 136 images in venous-arterial category, and 49 diabetic images. After pre-processing and removing some of the images, we ended up with 110, 110, and 45 images in each class, respectively. The removed samples included images in which the wound area was totally or partially covered by the bandage, noisy images, duplicate images, etc. 15% of the images were selected randomly as test set and the remaining images were used for training. Table 4.16 displays the results of applying our proposed model on this dataset. As we see in the table, for both binary and 3-class classification experiments, our proposed classifier shows a superior performance in comparison with the classifier A and B. Since the size of the images in Medetec dataset was larger than ours, we splitted the input images into different number of slices during sliding window part of the classifier B, to find the optimum number of slices. The best results obtained when we used a four by four grid in sliding window technique. We can conclude that for larger images we need to cut them into more slices when we use sliding window method, while for the smaller images, the best accuracy can be achieved using a lower number of slices.

Table 4.15: Performance comparison of different classifiers on our own dataset. All values are in percentage.

Classifier	Binary classification			3-class classification		
	Avg	Max	Training time	Avg	Max	Training time
ResNet101	90.35	94.60	~50 min	84.88	88.40	~75 min
ResNet50	85.34	92.86	~17 min	80.94	82.60	~26 min
SqueezeNet	78.58	89.30	~5 min	70.00	76.70	~6 min
Our ensemble classifier	**94.28**	**96.40**	~20 min	**87.70**	**91.90**	~28 min

Table 4.16: Medetec dataset results. All values are in percentage.

Classifier	Binary classification	3-class classification
A	82.40	75.60
B	73.50	63.40
Our ensemble classifier	**91.20**	**82.9**

4.6 Conclusion

Acute and chronic wounds are a challenge and burden to healthcare systems in all countries. The wound diagnosis and treatment process can be facilitated using an efficient classification method. Machine learning and deep learning have a good potential to be used as powerful algorithms for wound image analysis tasks such as classification. Prior works in the literature mainly dealt with binary classification or studied only specific types of wounds like diabetic ulcers. Additionally, the major part of the previous studies classified extracted ROIs or wound patches, rather than the whole wound images. Also, most of them had difficulties accessing valid,

reliable, and high-quality images as some of them collected their data from the web. To fill these gaps, we proposed an end-to-end novel ensemble deep learning-based classification method for classifying the chronic wounds into multi categories based on their type.

In this chapter, we proposed an end-to-end ensemble deep CNN-based classifier for classification of wound images into multiple classes based on the type of the wound. To the best of our knowledge, our proposed classifier is the first model that classifies the wound images into more than two types. We initially designed patch classifiers with fine-tuned AlexNet architecture to efficiently classify the wound patches into different wound types. Influence of different wound types on the classification accuracy was investigated by running numerous experiments and testing different combinations of the wound types. For image-wise classification task, for each input image, first a feature vector created using the designed patch classifier and another AlexNet that was trained on the whole images. Then the feature vector fed into an MLP to obtain an ensemble image-wise classifier with a higher accuracy and better performance. The obtained resutls showed that the combination of the two classifiers, resulted in a more strong classifier with a superior performance that outperforms each of the combined classifiers individually. This outcome confirms our hypothesis that combination of both patch level and whole image level information generates better classification results. Also, the results show that our proposed ensemble classifier can be used successfully as a decision support system in wound image classification tasks to assist the physicians in related clinical applications. We have made available the dataset we used for the current research.

4.6.1 Future directions

As a future study, we plan to improve the performance and classification accuracy of our proposed classifier by trying different combinations of the patch-wise and image-wise classifiers. Besides, testing the proposed approach to classify the images into more classes by working on a larger dataset of wound images would be one of the subsequent steps of this research.

Chapter 5

Wound Tissue Classification

In this chapter, we have a quick review on some of the researches in the literature in the field wound image tissue classification and analysis using traditional machine learning-based methods or deep convolutional neural networks. At the end, we discuss the gaps in this field as well as the weaknesses of the reviewed works as clues for our future steps in this thesis.

5.1   Related Works

In this section, we review the recent studies in wound tissue analysis field from the literature. We have organized the reviewed works under two main subbranches: superpixel generation-based methods and approaches that are based on analysis of the whole image. Each cagetory has been splitted into smaller groups considering the method that was used in the articles. Figure 5.1 displays the categorization chart for the reviewed papers.
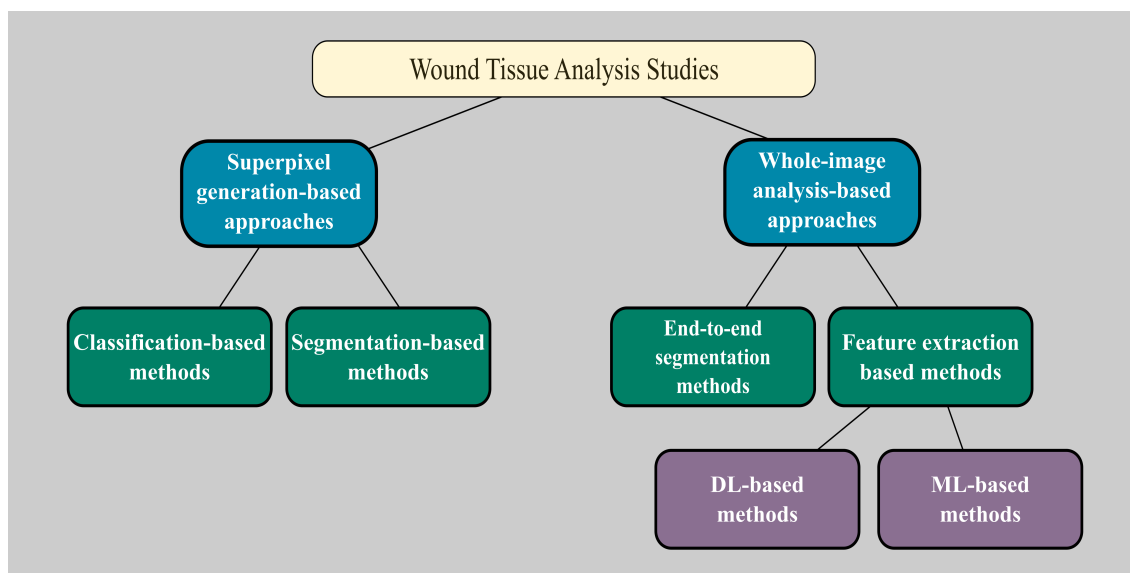
Figure 5.1: Organization chart for the wound tissue analysis papers.

5.1.1 Superpixel generation-based approaches

Classification-based methods

Zahia et al. proposed a method for tissue analysis of pressure wound images using deep CNNs [63]. The dataset they used in this research includes 22 pressure wound images. For each image, the ground truth labels for pixels specified by specialists. As a pre-processing task, the authors extracted the ROI from the original wound image and then removed the flash light from the images as well as extracting 5×5 patches from each one. At the next step, these patches along with their ground truth labels, fed into the CNN for training the network. The output of the network would be one of the labels granular, slough, and necrotic that shows the wound tissue type of the input image. In detail, during test phase, the input image is partitioned into several 5×5 patches. By classifying each patch, the whole segmented image will be generated.

69

The CNN used in this study is a 9-layer network with 3 convolutional layer which proposed by the authors. The authors used classification accuracy, Dice similarity coefficient, precission per class, sensitivity, and specificity metrics for reporting their results. Based on the results reported in the paper, the slough tissue is the hardest tissue to classify.

In another research, Rajathi et al. proposed a CNN-based method for tissue classification of varicose ulcer wound images [60]. The dataset used in this study included around 1250 varicose ulcer images collected by a medical college in India and a web page. Also, the images was leballed by specialists. Their method is similar to the approach used in [63]. Their approach includes three phases: data pre-processing, active contour segmentation for departing the wound area from the skin, and using a CNN to classify the input image. In pre-processing step they removed the flash light from the images using some simple image processing techniques like thresholding. In the second step, using the active contour segmentation method the ROI was extracted. In the final step, they used a CNN with 4 layers for classification that proposed by the authors. They generated 5×5 patches from the ground truth and the segmented wound image to feed into the network for training. The output would be the segmented wound image in which four different tissues of the wound have been classified: Granulation, Slough, Epithelial, and Necrotic.

In a recently published paper, G. Blanco et al. [65] proposed a method for dermatological wound image analysis named it QTDU that is a combination of deep learning models and superpixel-driven segmentation methods. In this study they used a dataset with 217 arterial and venous wound images from lower limbs. The

method proposed in this research includes three main steps. The first step is data pre-paration and region segmentation that includes labeling the images, constructing the superpixels, and augmentation of the superpixels. In this step they selected 40 images out of the total number of the images and extracted superpixels from each image. The superpixels were selected in a way to have 550 pixels each one. Finally, they provided 44893 superpixels and labeled them in a way that each one belongs to one of these classes: fibrin, granulation, necrosis, and and not wound. The second step is data processing and training stage in which two deep CNNs, Resnet and Inceptionv3 were trained by the data samples as well as 6 additional layers added to the end of these deep architectures. The output of the final layer is the label. Doing some experiments they also consluded that networks which have been pre-trained on ImageNet dataset were trained faster in compare to the architectures with random weights. In the final step, the pixelwise wound quantification masks are generated. In addition, the authors said that their proposed method generated better results when they used Resnet architecture in comparison with the case they utilized Inceptionv3. Also, testing the proposed method on 179572 superpixels, the results showed the QTDU method outperformed the machine learning approaches.

Segmentation-based methods

In another research, Niri et al. proposed a method for tissue classification of diabetic ulcers using superpixel generation and fully convolutional neural networks [39]. The goal was to classify the wound pixels into three different tissue types including necrotic, granulation, and slough. The method consists of two stages: automatic

wound segmentation and tissue classification. The dataset used in this study included 219 ulcer images which were captured using smartphone camera. At the first stage, wound segmentation was performed using U-Net architecture. Then, some morphological operations followed by ROI extraction phase, generated the desired wound area to be used in the next step. At the second stage, superpixels were generated using SLIC algorithm from the ROIs. Then 5000 superpixels along with their corresponding labels were utilized for training fully convolutional network like FCN-32. The proposed method was compared with some fully convolutional networks as well as SegNet and U-Net architectures and outperformed all of them in both accuracy and DICE index. The FCN-32 network that was trained using superpixels, showed the best performance among all networks.

5.1.2 Whole image analysis-based approaches

Feature extraction-based methods

ML-based methods   In 2015, F. Veredas and the coauthors published a paper in which proposed a machine learning based approach for wound segmentation and classification of the tissue in pixel level [38]. In more detail, they used k-means clustering method for the segmentation part and then utilized three different classifiers including neural networks, SVM, and random forest decision trees for the classification part. They used a dataset including 113 pressure wound images. After preprossing and segmentation, the researchers extracted some features from the segmented areas of the wounds to feed into the classifiers. Before training the classifiers, some experts labeled the segmented wound-bed and peri-ulcer regions using four and two labels

respectively. They reported high accuracy rates for the classifiers specifically SVM and random forest trees. Also, the lowest accuracy belong to neural network in this study.

DL-based methods   In another study, H. Nejati et al. proposed a deep learning based method to analysis chronic wound tissue and classifying it into one of the seven described classes mentioned in their paper [93]. Based on their claim, this paper is the first article for classifying wound tissue into more than four classes. The authors said that the previously published papers in the literature for wound tissue analysis can not be used for the clinical application because of considering only three types of tissue in chronic wounds. In this study, they used a pretrained AlexNet architecture for feature extraction from the wound tissue and then fed them into an SVM patch-level classifier for classification. The dataset they used in this study includes only 350 images. This is one of the limitations in their work. Also, the authors said that having an scarce dataset was one of the reasons that they couldn't use transfer learning to improve the network's performance. For the feature extraction part, they used two traditional feature extractors as well as AlexNet for comparison. Based on the reported results in the paper, the best results obtained when they used AlexNet along with the SVM classifier. The highest accuracy was related to the Infected category (95.54%) and the total accuracy is 86.40%.

In another study, Zapirain et al. utilized 3D convolutional neural networks for tissue classification of pressure wounds [94]. The goal in this research was to classify the wound pixels in the input image into one of the three tissue types: necrotic, granulation, and slough. The utilized dataset included 193 pressure wound images.

The first step was ROI extraction in which HSI color images obtained from the original RGB images along with gaussian smoothing of the original images were used as the input of a 3D DCNN to extract the ROI. At the second step, the extracted ROI along with LCDG model of the original image, HSI image, GS image, and a prior model are fed into another 3D DCNN for wound tissue segmentation. The 3D CNNs include 4 and 8 convolutional neural networks, respectively. After the convolutional layers there are two fully connected layers in both architecture. The authors compared their proposed method with some segmentation approaches in the literature and claimed that the results were promising.

End-to-end segmentation methods

In another article, Wang et al. proposed a method for wound region segmentation using deep convolutional neural networks [58]. Also, they used the features generated by the CNN for infection detection and healing status prediction. For the segmentation task a convolutional encoder-decoder used in this study. Also, for the infection detection part they designed a binary classifier in which fed the extracted features by the ConvNet into a SVM classifier. The output of the SVM classifier shows if the wound is infected or not. Finally, in the third part by having the previous wound surface area over the past N time units and the present image, they used a Gaussian process regression to predict the wound area for the future. Using this prediction they can estimate the healing time of the wound. The authors used the NYU Wound Database in their paper which includes more than 8000 wound images as well as some useful medical records like clinic visit dates and wound surface area

74

for each patient. Also, the researchers preprocessed the dataset images by cropping them into smaller sizes. In addition, they created wound segment annotations as well as binary infection labels to the images. To make the training easier and faster they utilized an NVIDIA Telsa K40 GPU in their research.

In [35] the authors proposed a deep learning based method for burn depth prognosis. Also, in this study they proposed a new wound dataset including 929 images most of them obtained from a medical center and the rest collected from the internet. For each image the burn area specified by three surgeons. Then each image labeled by three randomly selected surgeons. They defined four labels for the dataset images: superficial, superficial/deep partial thickness, full thickness, and undebrided. In the first step of their study, the authors classified the images into burn or no burn categories. The second step is related to separating the wound area from the rest of the image. In the final step they classify the burn wound into different classes which show depth levels of each one. The authors used a FCN (Fully Convolutional Network) for the network proposed in the paper. Also, they said that this network is based on the VGG-16 network. The authors used pixel accuracy and mean intersection over-union (IOU) as the metrics for evaluating the network performance of segmentation.

Godeiro et al. proposed an approach in which CNNs were used for wound tissue classification [66]. The goal was to classify the wound tissue into one of the three classes necrotic, granulation, and slough. The utilized dataset included 30 wound images that were captured by different smartphones and iPhone camera and were splitted into the three classes mentioned earlier. The proposed method consisted of four steps: preprocessing and noise reduction, wound segmentation, color

75

space reduction, and tissue classification using a DCNN. After noice reduction, a traditional image processing technique, Watershed algorithm, was used for segmenting the wound area. Then, the images were converted from RGB color space into CIELab domain. Finally, using different deep convolutional networks including U-Net, SegNet, FCN8, and FCN32, the method was evaluated. U-Net along with the color space reduction approach, generated the best classification results.

In another research, Goyal et al. used fully convolutional neural networks for semantic segmentation of skin lesions [95]. In more detail, and end-to-end multiclass semantic segmentation approach presented to segment three different skin lesions including melanoma, seborrhoeic keratosis, and naevus. The authors used a two-tier transfer learning method along with a hybrid loss function for training the deep architectures. A publicly available dataset including 2750 dermoscopy images was used in this study. Different FCNs including FCN-AlexNet, FCN-32s, FCN-16s, and FCN-8s were used for comparison. At the final step, a priority based post-processing algorithm was utilized to assign a single label to the detected lesion. FCN-8s generated the best outcome for the first two classes, while FCN-AlexNet showed the better performance for naevi category.

In another study, Goyal et al. proposed a method for diabetic foot ulcer segmentation using fully convolutional neural networks [96]. The authors presented a new dataset of foot images including 705 images which were captured by a Nikon camera. The goal in this research was to segment the wound and its surrounding skin automatically using FCNs. A two-tier transfer learning strategy was utilized for training of the deep networks. In this transfer learning strategy, first the network

76

was trained on ImageNet dataset and the first five convolutional layers transfered to the second step. Then, the network was trained on another large dataset, Pascal VOC. After pretraining on this dataset, all the layers transfered to the third step which was fine-tuning using the new wound image dataset. Different varients of FCNs such as FCN-AlexNet, FCN-32s, FCN-16s, and FCN-8s were utilized in this research. FCN-32s and FCN-16s showed a better performance in comparison with the other networks.

Wagh et al. compared a deep learning approach with a non-deep learning method for semantic segmentation of smartphone wound images [97]. In more detail, in the first strategy which is based on AHRF algorithm, they extracted features from the images and used traditional machine learning tools for classification. The AHRF approach considers the neighborhood pixels of a target pixel to assign it to the correct class. In the second approach, an end-to-end DCNN-based method was utilized for ulcer semantic segmentation. The dataset includes 1758 wound images from three different resources. Three different deep structures including FCN, U-Net, and DeepLabV3 were utilized in this study. The conclusion was that for datasets with less than 300 images, the AHRF based method showed better outcome in comparison with the U-Net architecture. The authors claimed that for larger datasets, all of the deep networks outperform the AHRF method.

In another research, Pholberdee et al. proposed a method which combines image processing techniques with deep learning methods for wound tissue semantic segmentation [98]. The proposed approach was utilized for semantic segmentation of the wound pixels into granulation, necrosis, and slough tissue types. Regarding the

obtained results, slough tissue was the most difficult tissue type for classification. The proposed convolutional network in this study, includes three convolution layers followed by two fully connected layers. Also, a color augmentation method was utilized for improving the segmentation accuracy. In comparison with an end-to-end segmentation method from the literature, the authors claimed that their proposed method generated better performance. This result was interesting as the structure of the deep network that was used in this study was much simpler than the previous work from the literature.

5.2   Dataset

For this project we obtained 110 images from different types of wounds collected by the AZH Wound and Vascular Center in Milwaukee. The images were captured using an iPad Pro (software version 13.4.1) and a Canon SX 620 HS digital camera. Most of them have the size of 320×427 and a few images have the size of 320×240 and all of them are in jpg format. The groundtruth was provided by three wound specialists from the clinic. The images were labeled in a way that each tissue type was identified with a specific color. Figure 5.2 shows sample images from the dataset beside their corresponding groundtruth. In these images, colors brown, red, yellow, green, and blue stand for callous and maceration, granulation, fibrin and slough, necrotic, and neodermis tissues, respectively.

Figure 5.2: Some of the samples from our dataset. The top row shows the original images and the bottom row displays their corresponding groundtruth.

5.3 Method

This section describes our wound tissue classification method. Figure 5.3 shows an overview of the whole process. Looking at Figure 5.1, our method goes under the superpixel generation-based approaches, as we splitt the input image into superpixels and then feed them into the DCNN-based classifier. Next subsections, describe each step in detail.

5.3.1 Preprocessing

ROI extraction

After investigaion of the dataset images, we removed some images because of some quality related issues like having very small size of the wound in the original image. Then we ended up with 73 images. At the next step, the Region of Interest (ROI)
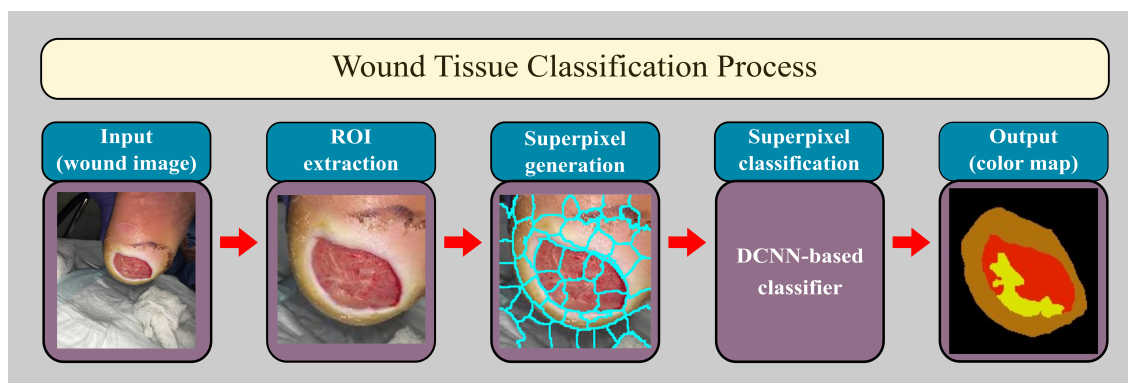
Figure 5.3: Overview for the wound tissue classification project.

was extracted manualy from each of the images.

Data splitting

In this step, we splitted the extracted ROIs into train and test sets, randomly. The train and test sets included 59 and 14 images, respectively. None of the training samples was used in the test set and vice versa.

Superpixel generation

After splitting the data samples, superpixels were generated from each ROI. For each superpixel, we defined the label using majoriy voting on the pixels in the corresponding region of the groundtruth. Simple Linear Iterative Clustering (SLIC) method was used for superpixel generation. The number of superpixels was identified based on the size of ROI. For larger ROIs, we increased the number of extracted superpixls.

Data augmentation

In data augmentation step, 16 samples generated from each superpixel in the training set using image transformation methods like rotating, flipping, cropping, and mirroring. Therefore, the number of superpixels that will be used for training of the classifier increases significantly and prevents the overfitting problem.

5.3.2 Training a DCNN

In this step, the extracted superpixels along with their labels are fed into a Deep Convolutional Neural Network (DCNN) for training. We used GoogLeNet architectures as one of the most common deep architectures that generated outstanding outcomes in different research area. We fine-tuned the network which was pretrained on ImageNet dataset, using the superpixels extracted from our own dataset. The maximum number of epochs was set to 20, the learning rate value was 1e-4, and the Adam optimization algorithm was used for training the network.

5.4 Results and discussion

In this research, MATLAB 2019b and 2020b versions were used for implementing the models. After training the model, it was applied on the test set images. Figure 5.4 shows some of the primitive results we obtained by applying the trained network on the test images. In this figure, the red collor shows granulation tissue, the yellow color displays the fibrin tissue, and the brown color shows callous tissue. As can be found from these primitive outcomes, the network couldn't detect the superpixels'

label in an efficient way and the results are not good visually. The justification for this outcome could be the fact that regarding the size of the images, the generated superpixels have very low resolution and are not informative to train the DCNN very well. We expect that by increasing the number of training samples and resolution of the original images, the issue would be solved.
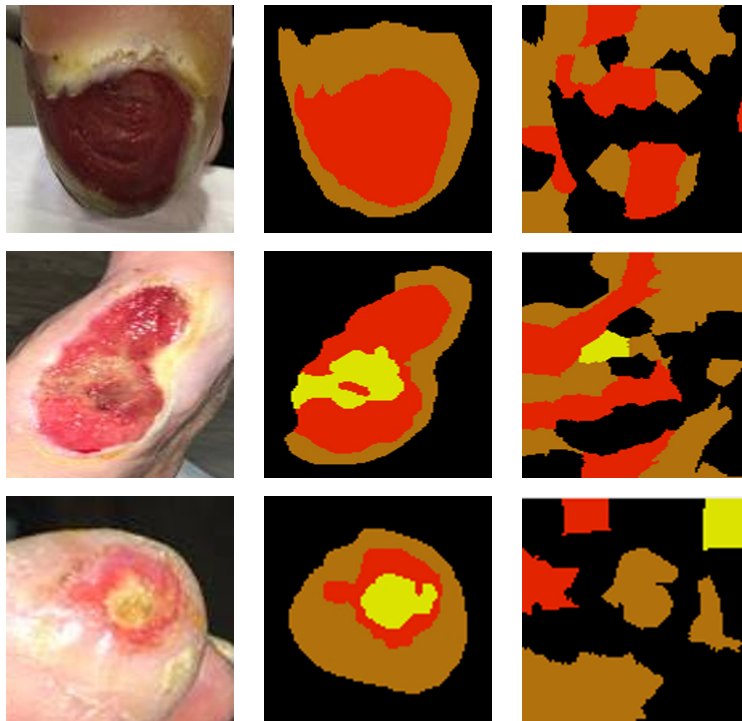
Figure 5.4: Some of the primitive results obtained from applying the trained model on the test set images. Each row shows the original images along with its corresponding groundtruth and output.

In this chapter, we reviewed some recent studies in the field wound tissue analysis that most of them used deep neural networks in their works. Also, a taxonomy was presented for the reviewed papers. By looking at the papers reviewed in the Section 5.1, it can be realized that there are some gaps in this area in the literature

as well as some weaknesses for the papers published in the field. First of all, in most of these reviewed papers the researchers worked on only a specific type of wound like diabetic, pressure, varicose and so on. The principal reason for this issue could be the fact that they did not access to a large and reliable dataset of different types of wounds. In addition, most of them classified the wound tissue to four classses and only one of them categorized the tissue into more than 4 categories. As a final note, in the reviewed works the researchers used AlexNet, Inceptionv3, ResNet, and VGG-16 networks or modified versions of these architectures. Filling these gaps in the literature was our main motivation for this research. In the future, we will collect more data samples and continue our experiments using different DCNNs to classify the wound tissues into more than four tissue types.

Chapter 6

Conclusion

6.1 Conclusion

In this thesis we studied medical image analysis task using convolutional neural networks which are one of the most common shapes of deep learning models. In chapter 2, we introduced deep learning in brief and discussed some of the most common DCNN architectures in detail. Also, we realized how useful deep learning models are in healthcare and wound care. In chapter 3, we studied burn wound image classification problem and used a deep convolutional neural network for classifying the burn wound images into two and three classes. For the binary classification case, our classifier showed a considerable improvement over the previous work in the literature. Moreover, despite of having a limited number of images we obtained an acceptable accuracy level for the 3-class classification problem. In chapter 4, we discussed multiclass wound image classification task using deep convolutional neural networks. At the first part of this chapter, we reviewed the existing wound image classification studies extensively and provided a comprehensive taxonomy for the reviewed papers. Then, as the main contribution of this thesis, we proposed an ensemble deep convolutional neural network-based classifier for classification of wound images into multiple categories based on the wound type. The ensemble classifier included two separate

classifiers worked in parallel on the same input. The output of these two classifiers combined to generate a higher classification accuracy. The proposed classifier was constructed based on this hypothesis that combining a patch level and whole image level information may result in a better classification performance. The performed experiments confirmed the hypothesis as the ensemble classifier outperformed both classifiers individually. Various experiments were conducted to find the best patch classifiers and whole image classifiers and then we constructed the ensemble classifier using a path-wise and image-wise classification strategy. The ensemble classifier's performance was tested using several experiments and the resutls outperformed some of the most common deep architetures. Moreover, its performance was further evaluated by testing on a publically wound image dataset. In addition, we introduced a new valuable wound image dataset collected by AZH Wound and Vascular Center in Milwaukee. Collaborating with the AZH Center, we labeled and processed the images to make them ready to use. We have already published the dataset online and it can be accessed using the link mentioned in Chapter 4. We claim that our proposed classifier can be used successfully for the wound image classification tasks or similar applications to help the physicians and wound specialists as a decision support system. Chapter 5, was about wound tissue classification using the superpixel generatin method and DCNNs. In this part, we first reviewed some of the recently published papers in the field and then described our method for classifying wound pixels into different tissues along with some of initial results.

6.2 Future Directions

The future directions for this thesis are:

- Completing labeling and pre-processing task of the high-resolution AZH Dataset images for the wound tissue analysis project.

- Classifying the wound tissue pixels into multiple classes using different deep convolutional neural networks.

- Proposing an ensemble classification method for classification of wound tissue pixels into multiple classes.

- Developing and publishing the AZH Wound Center Database. Doing this item is highly dependent on getting the required permissions from the AZH Wound and Vascular Center.

Bibliography

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in "Advances in neural information processing systems," (2012), pp. 1097–1105.

[2] G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdzal, S. Turcotte, C. J. Pal, S. Kadoury, and A. Tang, "Deep learning: a primer for radiologists," Radiographics **37**, 2113–2131 (2017).

[3] K.-H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," Nature biomedical engineering **2**, 719–731 (2018).

[4] P. Lakhani, A. B. Prater, R. K. Hutson, K. P. Andriole, K. J. Dreyer, J. Morey, L. M. Prevedello, T. J. Clark, J. R. Geis, J. N. Itri *et al.*, "Machine learning in radiology: applications beyond image interpretation," Journal of the American College of Radiology **15**, 350–359 (2018).

[5] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science **349**, 255–260 (2015).

[6] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: past, present and future," Stroke and vascular neurology **2**, 230–243 (2017).

[7] W. A. Figgett, K. Monaghan, M. Ng, M. Alhamdoosh, E. Maraskovsky, N. J. Wilson, A. Y. Hoi, E. F. Morand, and F. Mackay, "Machine learning applied to whole-blood rna-sequencing data uncovers distinct subsets of patients with systemic lupus erythematosus," Clinical & Translational Immunology **8**, e01093 (2019).

[8] M. Andreatta, V. I. Jurtz, T. Kaever, A. Sette, B. Peters, and M. Nielsen, "Machine learning reveals a non-canonical mode of peptide binding to mhc class ii molecules," Immunology **152**, 255–264 (2017).

[9] M. G. Bari, C. Y. Ung, C. Zhang, S. Zhu, and H. Li, "Machine learning-assisted network inference approach to identify a new class of genes that coordinate the functionality of cancer networks," Scientific reports **7**, 1–13 (2017).

[10] S. F. Rahman, M. R. Olm, M. J. Morowitz, and J. F. Banfield, "Machine learning leveraging genomes from metagenomes identifies influential antibiotic resistance genes in the infant gut microbiome," MSystems **3**, e00123–17 (2018).

[11] O. Collier, V. Stoven, and J.-P. Vert, "Lotus: A single-and multitask machine learning algorithm for the prediction of cancer driver genes," PLoS computational biology **15**, e1007381 (2019).

[12] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," Nature medicine **25**, 24–29 (2019).

[13] F. F. Ting, Y. J. Tan, and K. S. Sim, "Convolutional neural network improvement for breast cancer classification," Expert Systems with Applications **120**, 103 – 115 (2019).

[14] R. A. Castellino, "Computer aided detection (cad): an overview," Cancer Imaging **5**, 17 (2005).

[15] F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning," in "Proceedings of the 5th international conference on bioinformatics and computational biology," (ACM, 2017), pp. 59–62.

[16] C.-H. Liang, Y.-C. Liu, M.-T. Wu, F. Garcia-Castro, A. Alberich-Bayarri, and F.-Z. Wu, "Identifying pulmonary nodules or masses on chest radiography using deep learning: external validation and strategies to improve clinical practice," Clinical radiology (2019).

[17] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature **542**, 115 (2017).

[18] T. J. Brinker, A. Hekler, A. H. Enk, C. Berking, S. Haferkamp, A. Hauschild, M. Weichenthal, J. Klode, D. Schadendorf, T. Holland-Letz *et al.*, "Deep neural networks are superior to dermatologists in melanoma image classification," European Journal of Cancer **119**, 11–17 (2019).

[19] D. Y. Carson Lam, M. Guo, and T. Lindsey, "Automated detection of diabetic retinopathy using deep learning," AMIA Summits on Translational Science Proceedings **2018**, 147 (2018).

[20] Z. Gao, J. Li, J. Guo, Y. Chen, Z. Yi, and J. Zhong, "Diagnosis of diabetic retinopathy using deep neural networks," IEEE Access **7**, 3360–3370 (2019).

[21] R. Rostami, F. S. Bashiri, B. Rostami, and Z. Yu, "A survey on data-driven 3d shape descriptors," Computer Graphics Forum **38**, 356–393 (2019).

[22] G. Hinton, "Deep learning—a technology with the potential to transform health care," Jama **320**, 1101–1102 (2018).

[23] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" IEEE transactions on medical imaging **35**, 1299–1312 (2016).

[24] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," IEEE transactions on medical imaging **35**, 1285–1298 (2016).

[25] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering **22**, 1345–1359 (2010).

[26] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," arXiv.org (2018).

[27] S. Enoch and P. Price, "Cellular, molecular and biochemical differences in the pathophysiology of healing between acute wounds, chronic wounds and wounds in the aged," World Wide Wounds **13**, 1–17 (2004).

[28] P. Bowler, B. Duerden, and D. G. Armstrong, "Wound microbiology and associated approaches to wound management," Clinical microbiology reviews **14**, 244–269 (2001).

[29] T. N. Demidova-Rice, M. R. Hamblin, and I. M. Herman, "Acute and impaired wound healing: pathophysiology and current methods for drug delivery, part 1: normal and chronic wounds: biology, causes, and approaches to care," Advances in skin & wound care **25**, 304 (2012).

[30] C. K. Sen, G. M. Gordillo, S. Roy, R. Kirsner, L. Lambert, T. K. Hunt, F. Gottrup, G. C. Gurtner, and M. T. Longaker, "Human skin wounds: a major and snowballing threat to public health and the economy," Wound repair and regeneration **17**, 763–771 (2009).

[31] C. K. Sen, "Human wounds and its burden: an updated compendium of estimates," (2019).

[32] C. Lindholm and R. Searle, "Wound management for the 21st century: combining effectiveness and efficiency," International wound journal **13**, 5–15 (2016).

[33] G. A. James, E. Swogger, R. Wolcott, E. d. Pulcini, P. Secor, J. Sestrich, J. W. Costerton, and P. S. Stewart, "Biofilms in chronic wounds," Wound Repair and regeneration **16**, 37–44 (2008).

[34] M. Elmogy, B. García-Zapirain, C. Burns, A. Elmaghraby, and A. Ei-Baz, "Tissues classification for pressure ulcer images based on 3d convolutional neural network," in "2018 25th IEEE International Conference on Image Processing (ICIP)," (IEEE, 2018), pp. 3139–3143.

[35] O. Despo, S. Yeung, J. Jopling, B. Pridgen, C. Sheckter, S. Silberstein, L. Fei-Fei, and A. Milstein, "Burned: Towards efficient and accurate burn prognosis using deep learning," (2017).

[36] V. N. Shenoy, E. Foster, L. Aalami, B. Majeed, and O. Aalami, "Deepwound: Automated postoperative wound assessment and surgical site surveillance through convolutional neural networks," in "2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)," (IEEE, 2018), pp. 1017–1021.

[37] L. Russell, "The importance of wound documentation and classification," British Journal of Nursing **8**, 1342–1354 (1999).

[38] F. J. Veredas, R. M. Luque-Baena, F. J. Martín-Santos, J. C. Morilla-Herrera, and L. Morente, "Wound image evaluation with machine learning," Neurocomputing **164**, 112–122 (2015).

[39] R. Niri, H. Douzi, Y. Lucas, and S. Treuillet, "A superpixel-wise fully convolutional neural network approach for diabetic foot ulcer tissue classification," in "Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part I," (Springer International Publishing, 2021), pp. 308–320.

[40] D. Yadav, A. Sharma, M. Singh, and A. Goyal, "Feature extraction based machine learning for human burn diagnosis from burn images," IEEE Journal of Translational Engineering in Health and Medicine **7**, 1–7 (2019).

[41] M. Goyal, N. D. Reeves, S. Rajbhandari, and M. H. Yap, "Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices," IEEE journal of biomedical and health informatics **23**, 1730–1741 (2018).

[42] M. Goyal, N. D. Reeves, A. K. Davison, S. Rajbhandari, J. Spragg, and M. H. Yap, "Dfunet: Convolutional neural networks for diabetic foot ulcer classification," IEEE Transactions on Emerging Topics in Computational Intelligence (2018).

[43] C. Aguirre Nilsson and M. Velic, "Classification of ulcer images using convolutional neural networks," Master's thesis (2018).

[44] H. Alaskar, A. Hussain, N. Al-Aseem, P. Liatsis, and D. Al-Jumeily, "Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images," Sensors **19**, 1265 (2019).

[45] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1 (MIT press Cambridge, 2016).

[46] N. Ohura, R. Mitsuno, M. Sakisaka, Y. Terabe, Y. Morishige, A. Uchiyama, T. Okoshi, I. Shinji, and A. Takushima, "Convolutional neural networks for wound detection: the role of artificial intelligence in wound care," Journal of wound care **28**, S13–S24 (2019).

93

[47] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).

[48] "Imagenet large scale visual recognition challenge (ilsvrc)," `http://www.image-net.org/challenges/LSVRC/`. Accessed: 2019-10-31.

[49] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, "The history began from alexnet: A comprehensive survey on deep learning approaches," arXiv preprint arXiv:1803.01164 (2018).

[50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556 (2014).

[51] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in "Proceedings of the IEEE conference on computer vision and pattern recognition," (2015), pp. 1–9.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in "Proceedings of the IEEE conference on computer vision and pattern recognition," (2016), pp. 770–778.

[53] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in "Proceedings of the IEEE conference on computer vision and pattern recognition," (2016), pp. 2818–2826.

[54] O. A. Hiba Chougrad, Hamid Zouaki, "Convolutional neural networks for breast cancer screening: Transfer learning with exponential decay," in "31st conference on Neural Information Processing Systems (NIPS 2017)," (Long Beach, CA, USA, 2017).

[55] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," Annual review of biomedical engineering **19**, 221–248 (2017).

[56] F. Wang, L. P. Casalino, and D. Khullar, "Deep learning in medicine—promise, progress, and challenges," JAMA internal medicine **179**, 293–294 (2019).

[57] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," IEEE journal of biomedical and health informatics **21**, 4–21 (2016).

[58] C. Wang, X. Yan, M. Smith, K. Kochhar, M. Rubin, S. M. Warren, J. Wrobel, and H. Lee, "A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks," in "2015 37th annual international conference of the ieee engineering in medicine and biology society (EMBC)," (IEEE, 2015), pp. 2415–2418.

[59] F. Li, C. Wang, X. Liu, Y. Peng, and S. Jin, "A composite model of wound segmentation based on traditional methods and deep neural networks," Computational intelligence and neuroscience **2018** (2018).

[60] V. Rajathi, R. Bhavani, and G. Wiselin Jiji, "Varicose ulcer (c6) wound image tissue classification using multidimensional convolutional neural networks," The Imaging Science Journal pp. 1–11 (2019).

[61] M. Goyal, J. Ng, A. Oakley, and M. H. Yap, "Skin lesion boundary segmentation with fully automated deep extreme cut methods," in "Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging," , vol. 10953 (International Society for Optics and Photonics, 2019), vol. 10953, p. 109530Q.

[62] A. Abubakar and H. Ugail, "Discrimination of human skin burns using machine learning," in "Intelligent Computing-Proceedings of the Computing Conference," (Springer, 2019), pp. 641–647.

[63] S. Zahia, D. Sierra-Sosa, B. Garcia-Zapirain, and A. Elmaghraby, "Tissue classification and segmentation of pressure injuries using convolutional neural networks," Computer methods and programs in biomedicine **159**, 51–58 (2018).

[64] X. Zhao, Z. Liu, E. Agu, A. Wagh, S. Jain, C. Lindsay, B. Tulu, D. Strong, and J. Kan, "Fine-grained diabetic wound depth and granulation tissue amount assessment using bilinear convolutional neural network," IEEE Access **7**, 179151–179162 (2019).

[65] G. Blanco, A. J. Traina, C. Traina Jr, P. M. Azevedo-Marques, A. E. Jorge, D. de Oliveira, and M. V. Bedo, "A superpixel-driven deep learning approach for the analysis of dermatological wounds," Computer methods and programs in biomedicine **183**, 105079 (2020).

[66] V. Godeiro, J. S. Neto, B. Carvalho, B. Santana, J. Ferraz, and R. Gama, "Chronic wound tissue classification using convolutional networks and color space reduction," in "2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)," (IEEE, 2018), pp. 1–6.

[67] C. Chakraborty, "Computational approach for chronic wound tissue characterization," Informatics in Medicine Unlocked **17**, 100162 (2019).

[68] T. Fawcett, "An introduction to roc analysis," Pattern recognition letters **27**, 861–874 (2006).

[69] C. M. Bishop, *Pattern recognition and machine learning* (springer, 2006).

[70] J. Chauhan, R. Goswami, and P. Goyal, "Using deep learning to classify burnt body parts images for better burns diagnosis," in "Sipaim–Miccai Biomedical Workshop," (Springer, 2018), pp. 25–32.

[71] M. D. Cirillo, R. Mirdell, F. Sjöberg, and T. D. Pham, "Time-independent prediction of burn depth using deep convolutional neural networks," Journal of Burn Care & Research **40**, 857–863 (2019).

[72] A. Abubakar, H. Ugail, and A. M. Bukar, "Can machine learning be used to discriminate between burns and pressure ulcer?" in "Proceedings of SAI Intelligent Systems Conference," (Springer, 2019), pp. 870–880.

[73] A. Abubakar, H. Ugail, and A. M. Bukar, "Noninvasive assessment and classification of human skin burns using images of caucasian and african patients," Journal of Electronic Imaging **29**, 041002 (2019).

[74] "Biomedical image processing (bip) group from the signal theory and communications department (university of seville, spain) and virgen del rocío hospital (seville, spain)." `http://personal.us.es/rboloix/Burns_BIP_US_database.zip` (2020).

[75] "Imagenet," `http://www.image-net.org`. Accessed: 2019-11-11.

[76] M. Goyal, N. D. Reeves, S. Rajbhandari, N. Ahmad, C. Wang, and M. H. Yap, "Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques," Computers in Biology and Medicine p. 103616 (2020).

[77] L. Alzubaidi, M. A. Fadhel, S. R. Oleiwi, O. Al-Shamma, and J. Zhang, "Dfu_qutnet: diabetic foot ulcer classification using novel deep convolutional neural network," Multimedia Tools and Applications pp. 1–23 (2019).

[78] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical image analysis **42**, 60–88 (2017).

[79] K. K. Bressem, L. C. Adams, C. Erxleben, B. Hamm, S. M. Niehues, and J. L. Vahldiek, "Comparing different deep learning architectures for classification of chest radiographs," Scientific reports **10**, 1–16 (2020).

[80] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia, "Deep learning ensemble for hyperspectral image classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **12**, 1882–1897 (2019).

[81] S. Xia, Y. Xia, H. Yu, Q. Liu, Y. Luo, G. Wang, and Z. Chen, "Transferring ensemble representations using deep convolutional neural networks for small-scale image classification," IEEE Access **7**, 168175–168186 (2019).

[82] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Classification of histopathological biopsy images using ensemble of deep learning networks," arXiv preprint arXiv:1909.11870 (2019).

[83] B. Savelli, A. Bria, M. Molinara, C. Marrocco, and F. Tortorella, "A multi-context cnn ensemble for small lesion detection," Artificial Intelligence in Medicine **103**, 101749 (2020).

[84] D. Cha, C. Pae, S.-B. Seong, J. Y. Choi, and H.-J. Park, "Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database," EBioMedicine **45**, 606–614 (2019).

[85] E. Hussain, L. B. Mahanta, C. R. Das, and R. K. Talukdar, "A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network," Tissue and Cell **65**, 101347 (2020).

[86] D. Bermejo-Peláez, S. Y. Ash, G. R. Washko, R. S. J. Estépar, and M. J. Ledesma-Carbayo, "Classification of interstitial lung abnormality patterns with an ensemble of deep convolutional neural networks," Scientific reports **10**, 1–15 (2020).

[87] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).

[88] N. Van Noord and E. Postma, "Learning scale-variant and scale-invariant features for deep image classification," Pattern Recognition **61**, 583–592 (2017).

[89] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: Achievements and challenges," Journal of digital imaging **32**, 582–596 (2019).

[90] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," IEEE transactions on neural networks and learning systems **30**, 3212–3232 (2019).

[91] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," arXiv preprint arXiv:1602.07360 (2016).

[92] S. Thomas, "Medetec wound database [online]," (2021). `http://www.medetec.co.uk/files/medetec-image-databases.html`.

[93] H. Nejati, H. A. Ghazijahani, M. Abdollahzadeh, T. Malekzadeh, N.-M. Cheung, K.-H. Lee, and L.-L. Low, "Fine-grained wound tissue analysis using deep neural network," in "2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)," (IEEE, 2018), pp. 1010–1014.

[94] B. García-Zapirain, M. Elmogy, A. El-Baz, and A. S. Elmaghraby, "Classification of pressure ulcer tissues with 3d convolutional neural network," Medical & biological engineering & computing **56**, 2245–2258 (2018).

[95] M. Goyal, M. H. Yap, and S. Hassanpour, "Multi-class semantic segmentation of skin lesions via fully convolutional networks," arXiv preprint arXiv:1711.10449 (2017).

[96] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, and J. Spragg, "Fully convolutional networks for diabetic foot ulcer segmentation," in "2017 IEEE international conference on systems, man, and cybernetics (SMC)," (IEEE, 2017), pp. 618–623.

[97] A. Wagh, S. Jain, A. Mukherjee, E. Agu, P. C. Pedersen, D. Strong, B. Tulu, C. Lindsay, and Z. Liu, "Semantic segmentation of smartphone wound images: Comparative analysis of ahrf and cnn-based approaches," IEEE Access **8**, 181590–181604 (2020).

[98] N. Pholberdee, C. Pathompatai, and P. Taeprasartsit, "Study of chronic wound image segmentation: Impact of tissue type and color data augmentation," in "2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)," (IEEE, 2018), pp. 1–6.

Curriculum Vitae

Behrouz Rostami

EDUCATION

Doctor of Philosophy, Electrical Engineering

University of Wisconsin - Milwaukee, Milwaukee, WI   Spring 2021

Master of Science, Biomedical Engineering

K. N. Toosi University of Technology, Tehran, Iran   Feb. 2014

Bachelor of Science, Electrical Engineering

Lorestan University, Khoramabad, Iran   Feb. 2010

Associate Degree, Scientific-Applied Electrical Engineering

Hakim Sabzevari University, Sabzevar, Iran   Jul. 2007

INTERNSHIP EXPERIENCE

Post Graduate Intern   Oct. 2018-Aug. 2019

Model Predictive Maintenance (MPM) team, Advance Development Applications, Building Technologies and Solutions, **Johnson Controls, Inc.**, Milwaukee, WI, USA

Project: Design and develop AI Data-Driven model using RBF neural network for

mapping from degradation indexes to power model coefficients in VRF systems.

INDUSTRIAL EXPERIENCE

Technical support, maintenance and service engineer   2016-2017

IranBehdasht Company, Tehran, Iran.

Job description: Installation, fixing, and training Mortara devices at medical centers
and hospitals; After sales support services

Quality control engineer for EEG instruments   2015-2016

NegarAndishgan Company, Tehran, Iran.

Job description: Assembling, troubleshooting, packing, installation, and after sales
support services for EEG devices

RESEARCH EXPERIENCE

Research Assistant   Sep. 2017 - present

Department of Electrical Engineering, University of Wisconsin-Milwaukee, WI, USA.

Project: Wound image classification using deep convolutional neural networks

Research Assistant   Sept. 2012 - Feb. 2014

Department of Biomedical Engineering, K.N. Toosi University of Technology, Tehran,
Iran.

Project: Malignant and Benign Brain Tumor Detection Based on Type-2 Fuzzy
Systems

TEACHING EXPERIENCE

Graduate Teaching Assistant   Fall 2019 - Fall 2020

Department of Computer Science, University of Wisconsin-Milwaukee, WI, USA.

C Programming for embedded systems

Graduate Teaching Assistant   Fall 2018 - Spring 2019

Department of Computer Science, University of Wisconsin-Milwaukee, WI, USA.

Introduction to engineering programming (MATLAB)

Lecturer   Summer 2018

Department of Computer Science, University of Wisconsin-Milwaukee, WI, USA.

Introduction to engineering programming (MATLAB)

Graduate Teaching Assistant   Spring 2018

Department of Computer Science, University of Wisconsin-Milwaukee, WI, USA.

Introduction to engineering programming (MATLAB)

Lecturer   Summer 2014 - Fall 2015

Islamic Azad University of Shahryar, Tehran, Iran.

AVR micro-controllers (Assembly and C++), Electrical Circuits, Electronics LAB

JOURNALPUBLICATIONS

- <u>B. Rostami</u>, D.M. Anisuzzaman, C. Wang, S. Gopalakrishnan, J. Niezgoda, Z. Yu, "Multiclass Wound Image Classification using an Ensemble Deep CNN-based Classifier", Computers in Biology and Medicine (Under review).

- <u>B. Rostami</u>, S. Gopalakrishnan, J. Niezgoda, Z. Yu, "Multiclass Burn Wound Image Classification Using DeepConvolutional Neural Networks", Burns, (Under review).

- D.M. Anisuzzaman, C. Wang, <u>B. Rostami</u>, S. Gopalakrishnan, J. Niezgoda, Z. Yu, "Image-Based Artificial Intelligence in Wound Assessment: A Systematic Review" (Under review).

- R. Rostami, F. S. Bashiri, <u>B. Rostami</u>, Z. Yu, "A Survey on Data-Driven 3D Shape Descriptors", COMPUTER GRAPHICS *forum*, Vol. 38, No. 1, pp. 356-393, 2019 (Published).

- C. Wang, D.M. Anisuzzaman, V. Williamson, M. K. Dhar, <u>B. Rostami</u>, J. Niezgoda, S. Gopalakr- ishnan, Z. Yu, "Fully AutomaticWound Segmentation with Deep Convolutional Neural Networks", Scientific Reports, 10(1), 1-9, 2020 (Published).

- <u>B. Rostami</u>, M. Teshnehlab, "Gaussian Noise Reduction in Brain MR Images Using Fuzzy Hybrid Filters", 13th Iranian Conference on Fuzzy Systems , Jul. 2013 (Published).

- R. Rostami, F. FathNezhad, F. Aminzade, <u>B. Rostami</u>, R. Rostami, "An Arti cial Decision Support System base on Neural Networks to predict flood in a special geographic area", 1st National Conference on Software Engineering Applications, Iran, March 2009 (in Persian) (Published).

POSTER PRESENTATIONS

- B. Rostami, Z. Yu, "Breast Cancer Tumor Classification using Deep Convolutional Neural Networks", Poster Competition 2018, University of Wisconsin - Milwaukee, USA, April 2018.

- B. Rostami, Z. Yu, "Multiclass Wound Image Type Classification using Deep Convolutional Neural Networks", Poster Competition 2019, University of Wisconsin - Milwaukee, USA, April 2019.

HONORS AND AWARDS

Computer Graphics Forum's top 10% most downloaded papers   April 2020

WILEY.

Northwestern Mutual Data Science Institute (NMDSI) Scholarship-$5,000   Summer 2019

Northwestern Mutual Data Science Institute (NMDSI), Milwaukee, WI, USA.

Chancellors Graduate Student Award (CGSA)-$11,000   Fall 2017

Department of Electrical Engineering, University of Wisconsin-Milwaukee, WI, USA.

Chancellors Graduate Student Award (CGSA)-$2,000   Fall 2017

Department of Computer Science, University of Wisconsin-Milwaukee, WI, USA.