

FINE-GRAINED DIABETIC WOUND IMAGE ANALYSIS
AND AUTOMATED CLASSIFICATION SYSTEM USING
DEEP LEARNING

H.V.L.C.Gamage

188085B

Degree of Master of Science

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

January 2020

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgment any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other media. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 27-01-2020

Name: H.V.L.C.Gamage

The above candidate has carried out research for the MSc (Major Component of Research) Dissertation under my supervision.

Signature of the supervisor:.....

Date: 27-01-2020

Name of the supervisor: Dr. G. I. U. S. Perera

ABSTRACT

Diabetic ulcers are a major life-threatening complication among diabetic patients. The existing ulcer diagnosing practices depend on the visual examination of consultants. However, the precise manual diagnosing process is challenging since vision may vary upon the consultant, tedious and time-consuming. In the diagnosing process, the challenging task is to identify the infected areas and the severity of the ulcers.

Accordingly, automatic locating and segmenting of ulcer boundaries and severity stage classification is of significant prominence. Yet a comprehensive computer-aided Wagner scale based severity stage classification system for diabetic foot ulcers is not available in the literature. Even though there are few automated solutions for segmenting and locating of ulcer boundaries available in the literature, they consist of various limitations.

This research proposes solutions to automate two manual processes namely segmenting and locating ulcer boundaries and severity stage classification of diabetic ulcers. Here, a dataset of diabetic ulcers which consists of 2400 images was used for both tasks. Under the segmentation task, the process of instance-based diabetic ulcer segmentation was automated through the Mask-RCNN model. This solution could achieve 0.8605 of average precision value at 0.5 thresholds of Intersection over Union (IoU) and 0.5023 mAP value at 0.5 to 0.95 by the step size of 0.05 Intersection over union (IoU) threshold with ResNet-101 backbone for the DFU segmentation task.

In the meantime, an architecture to classify the severity stages of diabetic foot ulcers was implemented using DenseNet-201 pre-trained CNN architecture. In this approach, the classification head of the DenseNet-201 was removed and used the feature extraction head to extract the feature vectors. Then the feature reduction was done by applying a Global Average Pooling technique and used Singular Value Decomposition (SVD) as a further feature reduction technique. Additionally, SVD helps to optimize the memory consumption and processing time while preserving the accuracy of the proposed classification architecture. This proposed architecture could achieve an accuracy of over 96%.

ACKNOWLEDGMENT

First and foremost I extend my sincere gratitude towards my internal supervisor Dr. Indika Perera, of the Computer Science and Engineering, University of Moratuwa, for veteran assistance and guidance given throughout the research. Without his guidance and understanding, this research could not have been completed.

I would like to thank my external supervisor Dr. Manilka Sumanathilake for the guidance given at the beginning of the research.

To all the staff from the Department of Computer Science and Engineering who help me in multiple ways during carrying out this study, I extend my gratitude.

This research funded by the University of Moratuwa Senate Research Grant, and I extend my gratitude for the financial support provided by them.

I would like to express my heartfelt appreciation of the sacrifices made by my family members. I am grateful to them for their understanding and encouragement throughout the course.

To all who help and all my friends who made my time in the research lab delightful during the course.

TABLE OF CONTENT

DECLARATION	i
ABSTRACT.....	ii
ACKNOWLEDGMENT.....	iii
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xi
1. INTRODUCTION	1
1.1 Background	1
1.2 Motivation for the Research.....	4
1.3 Objective of the Research	5
1.4 Research Methodology.....	5
1.5 Publications	7
1.6 Thesis Orientation	8
2. LITERATURE REVIEW	9
2.1 Related Work.....	9
2.1.1 Image Segmentation.....	9
2.1.2 Conventional Medical Image Segmentation Approaches.....	9
2.1.3 Existing Software for DFU analysis	15

2.1.4	Recently proposed deep learning-based approaches for image segmentation	15
2.1.5	Image Classification Methods.....	20
2.2	Convolutional Neural Network	21
2.2.1	Feed-Forward Neural Network Vs Convolutional Neural Network...	22
2.2.2	Convolution Operation.....	23
2.2.3	Input Layer.....	24
2.2.4	Convolution Layer	24
2.2.5	Dropout Layer	24
2.2.6	Pooling Layer	25
2.2.7	Fully Connected Layer.....	26
2.2.8	Softmax Layer.....	26
2.3	Transfer Learning.....	26
3.	DATASETS	28
3.1	Diabetic Wound Dataset.....	28
3.1.1	Ground Truth Mask Generation.....	28
3.2	CVC-Colon Dataset.....	29
3.3	Chest X-ray Dataset	29
4.	METHODOLOGY	30
4.1	Task 1: DFU segmentation.....	30

4.1.1	Preprocessing	30
4.1.2	Model Architecture	30
4.1.3	Model Training	32
4.1.4	Hyperparameter Tuning	33
4.1.5	Reasons for choosing Mask R-CNN architecture	35
4.2	Task 2: DFU Classification	38
4.2.1	Preprocessing	38
4.2.2	Addressing the class imbalance problem	38
4.2.3	Approaches	39
5.	RESULTS & EXPERIMENTAL ANALYSIS	42
5.1	Results of DFU Segmentation	42
5.1.1	Evaluation Metrics	42
5.1.2	Comparison Model	43
5.1.3	Results on DFU Dataset	44
5.1.4	Results on Different Medical Datasets	46
5.2	Results of DFU classification	48
5.2.1	Evaluation Metrics	48
5.2.2	Results of Feature Extraction through CNN Models for DFU Classification	48
5.2.3	Results of Fine-tuning Pretrained CNN Models for DFU Image Classification	50

5.2.4	Comparison Models	50
5.2.5	Comparison Results on Proposed Approach and Comparison Models	51
6.	CONCLUSIONS.....	53
6.1	Contribution	53
6.2	Future work	54
	References.....	56

LIST OF TABLES

Table 1.1: Description of Wagner Scale	3
Table 4.1: Hyperparameter Table	33
Table 4.2 : Class Distribution of Training Dataset	38
Table 5.1: Results of DFU segmentation	45
Table 5.2: Results of different medical datasets	46
Table 5.3: Accuracy and F1-score of CNN feature extractor + ANN	49
Table 5.4: Accuracy and F1-score of CNN feature extractor + SVM	49
Table 5.5: Accuracy and F1-score of CNN feature extractor + Random Forest	49
Table 5.6: DFU dataset with SGD optimizer results	50
Table 5.7: Proposed approach with comparison models.....	52

LIST OF FIGURES

Figure 1.1: Visualization of Diabetes Complications in Human Body [1].....	1
Figure 1.2: Proliferation of Diabetes Mellitus [2].....	2
Figure 1.3: Visualization of Wagner Ulcer Grading Scale	3
Figure 2.1: Visualization of Differences in Classification, Object Detection, Semantic Segmentation, Instance Segmentation	9
Figure 2.2: Results of [14] Comparison.....	12
Figure 2.3: Architecture of Goyal et al. Proposed method [33].....	17
Figure 2.4: Architecture of Wang et al.s’ proposed system[34]	17
Figure 2.5: Architecture of Xiaohui et al s’ research work [35].....	18
Figure 2.6: Traditional CNN Architecture.....	22
Figure 2.7: Down Sampling	22
Figure 2.8: Initial Convolutional Operation.....	23
Figure 2.9: Convolution operation after moving by a stride.....	23
Figure 2.10: Visualisation of min, max, and average pooling	25
Figure 2.11: Global Average Pooling	26
Figure 3.1: User Interface of VGG Annotator Tool	28
Figure 3.2: Samples of Polyps dataset	29
Figure 3.3: Samples of Chest X-ray dataset.....	29
Figure 4.1: Mask R-CNN Architecture [6].	32

Figure 4.2: Workflow diagram of Mask R-CNN.....	34
Figure 4.3: R-CNN Architecture.....	35
Figure 4.4: Fast R-CNN Architecture	36
Figure 4.5: Architecture of Faster R-CNN	37
Figure 4.6: DFU Classification Architecture	41
Figure 5.1: Visualization of TP, FP, and FN	43
Figure 5.2: U-Net Architecture [59]	43
Figure 5.3: Few results of ulcer boundary segmentation task.....	45
Figure 5.4: Segmentation results for several neuropathic ulcers	46
Figure 5.5: Segmentation Results for Chest X-ray dataset with Mask R-CNN.....	47
Figure 5.6: Segmentation Results for Polyps dataset with Mask R-CNN	47

LIST OF ABBREVIATIONS

DFU	Diabetic Foot Ulcer
RPN	Regional Proposal Network
GAP	Global Average Pooling
SVD	Singular Value Decomposition
ROI	Region of Interest
AP	Average Precision
CNN	Convolutional Neural Network
IoU	Intersection over Union
SVM	Support Vector Machine
ANN	Artificial Neural Network
SGD	Stochastic Gradient Descent
RGB	Red Green Blue
HSI	Hue (H), Saturation (S), Intensity (I)
FC	Fully Connected
FCN	Fully Connected Network

1. INTRODUCTION

1.1 Background

Diabetes Mellitus can be introduced as one of the most pandemic and expansive chronic diseases [1]. The major intention behind this is the inability to utilize the insulin levels over a lengthy time period. Two different types of diabetes mellitus have defined namely type 1 and type 2. The inability of producing insulin is the leading cause of Type 1 diabetics and the type 2 diabetes effect due to the insulin resistance by the body cells or lack of insulin production by the pancreas. In the human body, insulin consent glucose to move from the blood into the liver, fat cells and muscles. Consequently, the imbalance of the quantity of insulin due to diabetic mellitus may lead to life-threatening complications such as cardiomyopathy, nephropathy, retinopathy, limb amputation, strokes, and neuropathy. Figure 1.1 indicates the complications of diabetes.

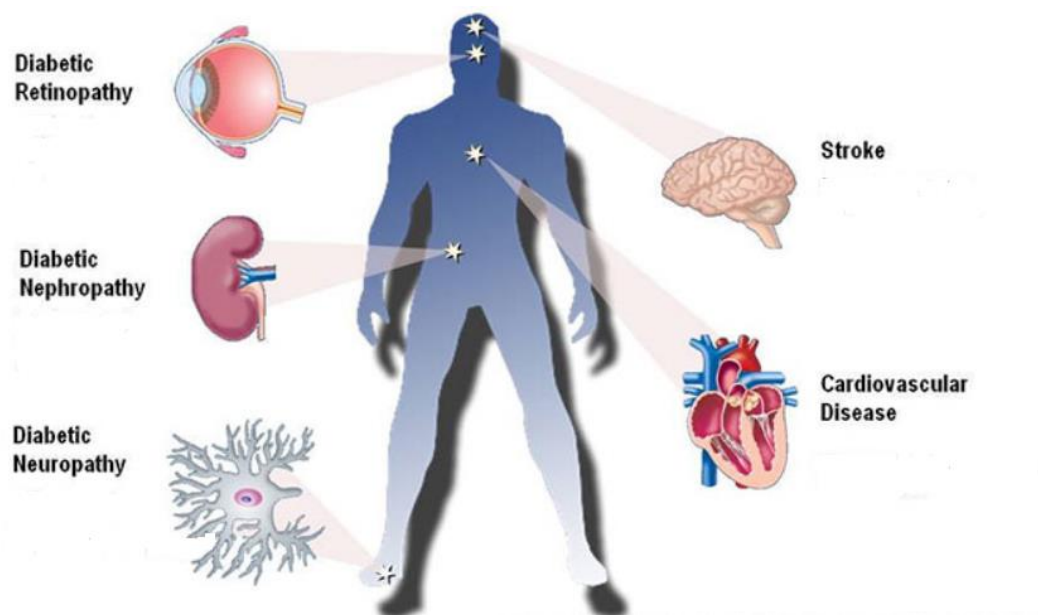


Figure 1.1: Visualization of Diabetes Complications in Human Body [1].

According to the latest estimation, 415 million individuals are experiencing diabetes mellitus comprehensively, representing 12% of the entire population, and however, 1 in 2 people stay undiscovered and untreated [2]. Figure 1.2 indicates the proliferation

of diabetes mellitus all over the world. Among the aforementioned complications of diabetic mellitus, Diabetic Neuropathy and Diabetic Foot Ulcers are the leading factors of limb amputation.



Figure 1.2: Proliferation of Diabetes Mellitus [2].

Diabetic Foot Ulcer (DFU) is an open wound on the patient's foot that does not heal on its own. Normally 15% of patients with diabetes will develop a diabetic foot ulcer in their lifetime [2]. There are different reasons for the proliferation of DFUs. Nerve damage is one of the reasons and as a result of this, the patient loses his/her feelings so the ulcers spread without any feelings. Changes in blood circulation are another reason and this leads to the development of ulcers and the problem of healing since it is harder for blood to reach the ulcer region. Deformation of feet is the other reason for occurring DFUs. The Wagner Scale is the most frequently used and stated as the gold standard in classifying DFUs. According to this scale, DFUs classify into 6 severity stages as described in Table 1.1 and decide treatments according to the severity.







					
Grade 0 No open Lesion	Grade 1 Superficial Lesion	Grade 2 Deep Ulcer	Grade 3 Abscess/ Osteomyelitis	Grade 4 Partial Foot Gangrene	Grade 5 Whole Foot Gangrene

Figure 1.3: Visualization of Wagner Ulcer Grading Scale

Table 1.1: Description of Wagner Scale

Grade	Description
0	Don't have open scratches but possibly will have cellulitis or deformity
1	Superficial diabetic ulcer(partial or full thickness)
2	Ulcer extension to ligament, tendon, joint capsule or deep fascia without abscess or osteomyelitis.
3	Deep ulcer with abscess, osteomyelitis or joint sepsis
4	Gangrene localized to a portion of forefoot or heel
5	Extensive gangrenous involvement of the entire foot

Nowadays DFU management raises various types of questions on patients and caregivers. For instance, routinely diagnosing and controlling with essential guidelines for patients in self-management are required to anticipate critical complications and diminish the possibility of life-long conditions.

The high amount of medical data that was collected during treatments has formed remarkable excitement in DFU care. Especially, image records afford a high influence to develop novel insights and disrupt the existing concerns on DFU care.

Computer vision-based technologies are dynamically robotizing recent medical practices and comprise higher precision, productivity, and fulfillment. The fast advancement of digitized information securing, computer vision and machine learning, and AI arrangements are beneficial for replacing the areas that were already done physically and thought to be as it were beneath the organization of human specialists.

1.2 Motivation for the Research

This section discusses the several factors that motivate the research undertaken in this MSc study. There is a dearth of research on diabetic foot ulcer classification through deep learning techniques.

When considering the current research arena, medical imaging broadly utilized for determination, treatment arranging, and surveying reaction to treatment in cutting edge pharmaceutical. As examples [3], [4] have used medical imaging for skin lesion analysis for melanoma detection, brain tumor classification and detection respectively.

In the current medical practices, DFU evaluation is performed by manual examination and using traditional measurements such as width, height, color, etc. The manual DFU examination procedure is a complex and tedious one so the remaining burden of the restorative specialist increments essentially because of the number of patients taking an interest in population screening. Moreover, this process consists of another drawback such as wrong decision making based upon the different vision of different consultants, inaccuracy on traditional measurements, difficulties in making measurements with patients who have widespread DFUs, etc.

Diagnostic decision-making has generally included utilizing proof from the information of the patient with the combination of the involvement of the doctors for comparable cases. Such similar cases are identified by the clinical staff primarily based upon visual properties.

Most of the diabetic patients are suffering from diabetic foot ulcers, but they do not care about DFUs at the early stages and do not make attention until the heavy growth of DFUs. The patients go untreated until later stages because they do not like to spend their time in long queues in hospitals. Such a patient can be remotely diagnosed and treated by these autonomous systems.

1.3 Objective of the Research

The purpose of this research is to investigate automatic approaches for diabetic foot ulcer detection and classification that can contribute towards improving diabetic foot management and subsequently, to develop an efficient system for diabetic foot ulcer analysis remotely. Basically, the proposed approach consists of two main tasks in DFU analysis as stated below.

1. To develop an automatic DFU detection, segmentation, and boundary screening system using DFU imagery in order to visualize the proliferation of DFU.
2. To develop an automatic DFU classification system in order to identify the most accurate severity stage of the ulcer.

1.4 Research Methodology

This study presents a novel automatic system for the segment and classification of diabetic foot ulcers. In this section, we briefly explain our methodology of the research work. As the initial step, we analyze the importance of the autonomous system for analyzing DFUs and its impact on the overall quality of diabetic care. Then we read on the related work conducted in the same disease (DFU) and the other different kinds of diseases. When studying their results and findings, we could identify the limitations. We identified the possible approaches to overcome those limitations and could obtain

more accurate solutions using the latest technologies. In order to achieve the objectives of this study, we implemented two automated approaches as summarized under Task 1 and Task 2. We use a dataset of 2400 ulcer imagery which is collected from DFU clinics in Sri Lanka and different online resources. We labeled the collected imagery with the help of a well-trained diabetic clinician. Moreover, we needed the corresponding annotations of DFUs for Task 1. Therefore we use an annotation tool [5] to do this and verify the annotated imagery with the assistance of the DFU clinician. Then we apply data pre-processing techniques to the dataset. After preprocessing, the dataset was divided into training and testing where the training set represents 80% and the test set represents 20% from the whole dataset. Then we applied a few data augmentation techniques for the training dataset.

Task 1: DFU Segmentation

For this automation task, we use an existing instance-based segmentation model and do the fine-tuning of the model to obtain the best results specific to our dataset. Initially, for this task, we use the Mask – RCNN [6] model with the COCO dataset [7]. At that point, we tweak the hyperparameters to improve the outcomes from the model for our dataset.

Task 2: DFU Classification

In task 2, we utilize a convolutional neural network engineered from the DenseNet 201[8] feature extractor paradigm together with a global average pooling (GAP) layer for the severity stage classification of diabetic foot ulcers. We use singular value decomposition (SVD) in order to reduce memory consumption and processing time while preserving the classification model performance.

1.5 Publications

The approaches described in this dissertation has been presented in international conferences. Below is a list of three publications arising during the course of this MSc study.

- **H. Gamage**, W. Wijesinghe, and I. Perera. Instance-based segmentation for boundary detection of neuropathic ulcers through Mask-RCNN. In International Conference on Artificial Neural Networks, pages 511–522. Springer, 2019.
DOI: https://doi.org/10.1007/978-3-030-30493-5_49
- **H. Gamage**, W. Wijesinghe, and I. Perera. Automatic Scoring of Diabetic foot Ulcers through Deep Convolutional Neural Network based Feature Extraction. In Bioinformatics and Bioengineering (BIBE) IEEE, 2019.
DOI: <https://doi.org/10.1109/BIBE.2019.00069>
- Isuru Wijesinghe, **Chathurika Gamage**, Indika Perera, Charith Chitraranjan, “A Smart Telemedicine System with Deep Learning to Manage Diabetic Retinopathy and Foot Ulcers”. In Mercon 2019. (Collaborative research work)
DOI: <https://doi.org/10.1109/mercon.2019.8818682>

1.6 Thesis Orientation

This work sectioned in different chapters:

Chapter 2 describes the different previous studies which are related to this study. Moreover, this presents an explanation of the theoretical backgrounds of the techniques that we use throughout this research study.

Chapter 3 is to describe the proposed architecture in order to address Task 1 & 2, with details about its parts, the methods, and the techniques that were espoused and combined. Furthermore, this chapter discusses the datasets, preprocessing steps, performance measures used to explore the given problem.

Chapter 4 focuses on the static analysis of methods, the feature selection process, the classification models and segmentation models.

Chapter 5 is devoted to results obtained from this research work and discussion on results as well as the challenges that we faced during this work and practical possibilities to overcome those.

Chapter 6 summarizes our findings and concludes the overall results by comparing it with other recently published works. Moreover, the final chapter discusses future research directions.

2. LITERATURE REVIEW

2.1 Related Work

2.1.1 Image Segmentation

In computer vision, there are different techniques available for image analysis namely object detection, classification, and segmentation. The classification utilized to distinguish the class of the object and it doesn't consider the detailed pixel-level structure of an image. Object detection utilized to distinguish both the class of the object and the spatial location. Segmentation use to partition an image into coherent parts. In the segmentation process, there are two ways namely semantic segmentation and instance segmentation. Semantic segmentation does the partition of the image into meaningful parts and does the classification of each part into pre-determined classes. Instance segmentation is very similar to the semantic segmentation, but this has an improved feature to identify each pixel, the object instance it belongs to.

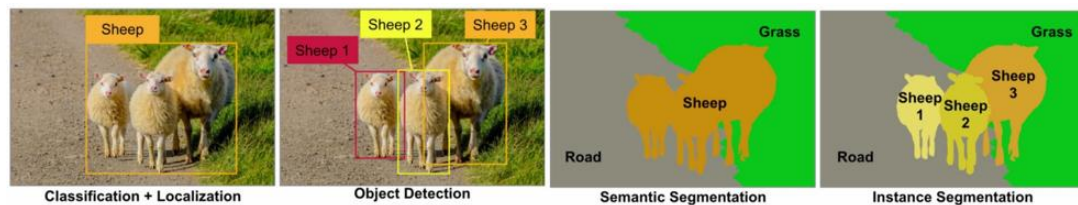


Figure 2.1: Visualization of Differences in Classification, Object Detection, Semantic Segmentation, Instance Segmentation

2.1.2 Conventional Medical Image Segmentation Approaches

In the recent past, different segmentation techniques appeared in the literature for medical image processing. Many schemes and techniques have been developed by the researchers in order to detect and segment the ulcer imagery. In addition to the ulcer segmentation, we study the methods that use in some other applications such as boundary detection in angiograms of coronary, tumor detection and segmentation, heart segmentation and analysis of cardiac imagery, etc. Here onward we describe the

different segmentation approaches beneath four categories specifically, region-based methods, classifier methods, clustering methods, and deep learning methods.

Region-Based Approaches

Region-based image segmentation algorithms depend on the properties of the region and boundary of the instance. Here we examine most usually utilized two region-based methodologies to be specific region growing and thresholding.

Region Growing

Region-based segmentation is a method which split an image into homogeneous areas of neighboring pixels using homogeneity criteria among candidate sets of pixels. Every pixel in a region is similar with respect to some characteristics such as intensity, texture, color, etc [9].

The Pseudocode of the algorithm is as below

Input (seed point)

Region $r = \{\text{seed point}\}$

While $r.\text{neighbors} \neq \{\}$

For each voxel x in $r.\text{neighbors}$, if $P(x,r) = \text{true}$

Then add x to r

End While

Return r

According to the pseudocode initially, one or more seed points should be given manually. Then the selected seed point(s) are considered as exact regions. The regions

are grown with adjacent points of selected seed point(s) depending upon the given criteria.

Perez et al. 2001 [10] propose an approach using the region's growing method for leg ulcer segmentation. The authors have considered the color property of the pixel to implement their approach. They have considered five color channels namely R, G, B, S and I from two color spaces of RGB and HSI. Among these five color channels, only one channel which performs better result has chosen for segmentation. The selecting best color channel has been the most challenging task to them because of the different conditions of the imagery, so they have used a specific algorithm for this. A hybrid approach was proposed by Veredas et al. [11] for the classification of wound tissues by combining the Bayesian classifiers and neural networks. They have used the region-growing based algorithm for wound image segmentation.

Filko.D et al. [12] have done a study on Detection, Reconstruction and Segmentation of Chronic Wounds Using Kinect v2 Sensor. At their pre-processing stage, they have used a region-growing based algorithm for wound image segmentation.

The drawbacks of this method are the outputs of this method completely depends on the selection of seed point. Generally, the Seed point(s) is selected by the human, so this method highly depends upon the human vision. This technique makes a gap within the extracted shape or delivered a detached region.

Thresholding

Thresholding is the simplest method in image segmentation and it is done based on an assumption that the images consist of regions with different gray levels [13]. İlkin et al. [14] have done a comparison of global histogram-based thresholding methods using an ulcer imagery dataset. They have considered five thresholding methods namely minimum error threshold, maximum entropy, simple threshold selection mean, simple threshold selection minimum, and Otsu's method for their comparison. In order to analyze the performance, they have considered MSE(Mean Square Error), TE(Time elapsed), PSNR(Peak Signal-to-Noise Ratio) and CI(Correlation Index) metrics.

Based on the results obtained by the authors, prove that the most successful method is maximum entropy methods and Otsu's thresholding for images with various properties. Figure 2.2 shows the results of the comparison. Fraiwan L et al. [15] have done a feasibility study for diabetic foot ulcer mobile detection system using a smartphone thermal camera. In their study, they have applied the Otsu thresholding method for image segmentation at their pre-processing stage. In addition to the wound segmentation, we could find different thresholding approaches for medical image segmentation namely MRI brain image segmentation [16], Cancer image segmentation using Fuzzy Entropy with Level Set Algorithm, Diacom image segmentation using fuzzy thresholding [17].

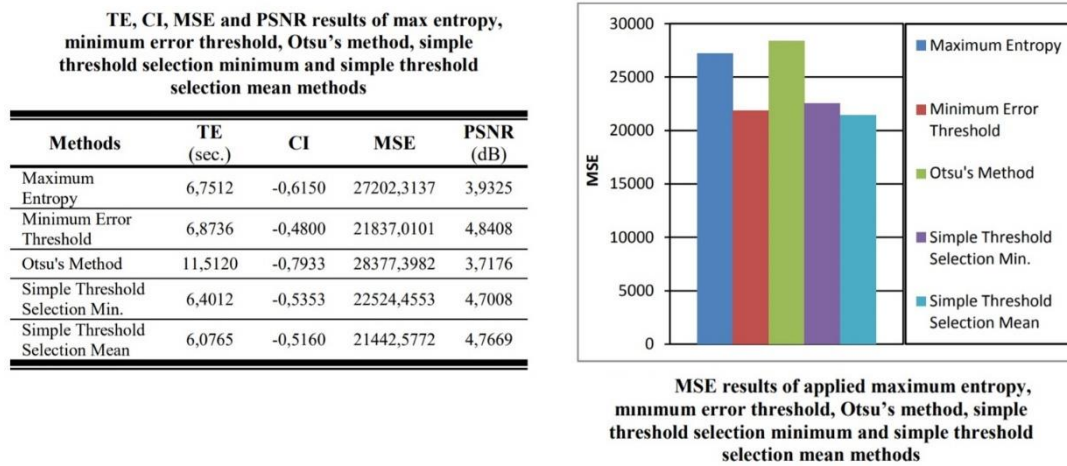


Figure 2.2: Results of [14] Comparison

Clustering Method

Yadav et al. [18] have used clustering techniques with selected color space for chronic wound segmentation. In their study data were collected from a normal digital camera and initially they have applied a combined gray world and retinex method for color correction. After then a 5 x 5 median filtering has used for noise reduction and anisotropic diffusion has used for homogenization. The authors have selected Dr and Db chrominance channels of YDbDr color space by considering the comparison of fifteen color spaces based on mean contrast between the wound and non-wound regions. Then they segment the wound regions using two clustering techniques namely

fuzzy c-means and k-means on Dr and Db channels. The obtained accuracy for the Dr Channel using k-means is 74.39% and for the fuzzy c-means is 72.55%. The accuracies for the Db channel are 73.76% with k-means and 75.23% with fuzzy c-means. As the final conclusions, they have proven that the fuzzy c-means algorithm by applying for the Db channel provided a higher accuracy for wound segmentation.

Paulo M. et al. [19] have done research work for dermatological ulcer segmentation using the clustering of color components. Initially, they have chosen HIS color space according to the previous literature. Therefore all the imagery in their dataset has converted into HIS color space. Then for each image, they have generated a hue-saturation (HS) histogram. For each generated histograms, they have applied a color clustering technique using an expectation-maximization procedure based on a Gaussian model. After the color clustering, the selection of a set of clusters that represent the composition of ulcer tissues has done manually. The final ulcer area has calculated using an automatically convex hull deriving method in order to fill the residual gaps. They have used 172 imagery as their test set and have obtained a 0.56 average Jaccard coefficient with a 0.22 standard deviation between computationally calculated ulcer area and the same region defined by a dermatologist manually.

Dhane et al [20] have proposed a novel method for wound area detection using optical imagery. Their approach consists of a spectral approach for clustering, based on the affinity matrix. This method includes the construction of a similarity matrix of Laplacian based on the Ng-jorden Weiss algorithm. Initially, they have pre-processed the imagery for color homogenization. The first-order statistics filter has applied to extract spurious regions and the filter has selected by considering the performance and evaluation of four quality metrics. After the segmentation of the imagery through a spectral method, the segmented regions have post-processed using morphological operators. They have used ground-truth imagery labeled by the dermatologists to evaluate the segmentation results. Furthermore, they have done a comparison of the segmentation results with the results of another two clustering methods namely k-means and Fuzzy C-means. The size of their test dataset is 105 images and have obtained the 86.73% accuracy and 89.54% sensitivity.

The drawbacks of the clustering approach are we cannot refer to special information, the number of clusters, the user should select the k value in the k-means algorithm, and sensitivity to outliers, noise, initial values.

Deformable Methods

Jones et al. [21] propose level set-based methods and active contour-based methods in order to adaptively regularize the segmentation contour upon the existing manual delineation process. This research contains limitations, in particular, failures at dealing with the false edges and poorly defined wound boundaries, sensitivity to illumination conditions, and depend on the initial contour settings.

Classification Methods

Kolesnik et al. [22] propose an approach to segment wound area in an image using a 3D color histogram-based SVM classifier. Their proposed method has shown a 25% increase in the performance compared to 5 segmentation methods namely Independent Sampling, Learning Vector Quantization, Vector Quantizer Design, Random Density Estimation, and Histogram Sampling.

In order to determine the boundaries of the foot ulcer images, a cascaded two-stage Support Vector Machine architecture was proposed by Wang et al. [23]. They use a k number of binary SVM classifiers. In the training process, they use a different subset of the entire training dataset. Misclassified instances were collected at the first stage of training and another SVM classifier was trained using these misclassified samples at the second stage of training. In the feature extraction process, they extract different features such as texture, color, etc using different image processing techniques. They have applied PCA to the extracted feature vector for further reduction of the size of the vector. Approached average sensitivity is 73.3% and average specificity is 94.6% through this architecture.

Veredas et al. [24] propose Bayesian classifiers and neural networks-based approach to do the binary tissue classification on ulcer images. In their research work, the region

segmentation has done using a clustering-based approach which proposes a hybrid model as the combination of a region growing strategy mean shift procedure.

2.1.3 Existing Software for DFU analysis

According to the literature, for wound image analysis and monitoring, we found several existing software systems. “PictZar Digital Planimetry” [25], Silhouette TM [26] by Aranz Medical, MOWA – Mobile Wound Analyzer[27], WITA by Filko et al. [28] are the examples for existing wound analysis software. All of the aforementioned systems have the ability to analyze the wound images, but they are still with the limitation of failure to automatic wound boundary detection. Therefore, the detection of wound boundaries is performed manually by system users.

2.1.4 Recently proposed deep learning-based approaches for image segmentation

There are different image segmentation algorithms available in the literature. In the past, rigid algorithms have been used in most cases, but they are less efficient compared to the deep learning algorithms because they heavily depend on human intervention and expertise. In section 2.1.2 I have described the usage and limitation of rigid algorithms for medical image segmentation. Deep learning architectures have been used in most of the modern image segmentation approaches because they require less human intervention and demonstrate a considerable performance.

This section describes the intuition behind the state of the art approaches which have been used for object detection and segmentation tasks and evolution from one implementation to the next. Girshick et al. [29] have introduced R-CNN architecture for semantic segmentation and object detection. Initially, R-CNN generates 2000 ROIs (Region of Interest) using a selective search algorithm. After then perform the classification through CNN for each and every ROI. Finally, it uses FC (Fully Connected) layers to classify and refine the bounding boxes. This architecture consists of drawbacks such as it’s training time consumption is very high since it has to classify 2000 ROIs per image, the selective search algorithm is a fixed algorithm so then we cannot verify the accuracy of a generation of ROIs.

Girshick et al. [30] have proposed the improved architecture of R-CNN called Fast-RCNN. In this architecture, they have feed the image directly to the CNN feature extractor and also they have externally used regional proposal methods in order to create ROIs. After then they have combined the feature maps and the corresponding ROIs to create the patches for object detection. ROI pooling has been used to wrap the patches into a fixed size. These patches have fed to the FC layers in order to perform the classification and the localization. This architecture faster than the [29] because no need to feed the 2000 ROIs to CNN every time and the CNN operation is carried out only once per image and a feature map is generated from it.

Girshick et al. [31] have proposed the improved architecture of Fast-RCNN called Faster-RCNN. This architecture is very similar to the [30], but this uses a separate network to predict ROIs instead of a fixed algorithm(selective search). This architecture is faster than the [29] & [30].

K. He et al. [32] have proposed a technique for instance based segmentation called Mask-RCNN. This architecture is based on Faster-RCNN architecture. Faster-RCNN architecture is a combination of a CNN backbone network and a regional proposal network (RPN). This model has two outputs namely an anchor box for the object region and the class label. Mask-RCNN provides an additional object mask as an output that supports to perform the object segmentation more precisely.

Goyal et al. [33] propose an automatic technique for diabetic wounds segmentation through deep learning. The fully convolutional network based on two-tier transfer learning has used in their proposed architecture. In the first tire, they have used ImageNet dataset in order to train the FCNs. In the second tire, the Pascal VOC segmentation dataset has used to train the FCNs. FCN-AlexNet, FCN-32s, FCN-16s, and FCN-8s have used as the different transfer learning models. FCN-16s has identified as the best performer out of these models. Moreover, the model has achieved the Dice Similarity Coefficient for ulcer region of 0.794, DSC for the surrounding skin region of 0.851, and DSC for the combination of both regions of 0.899 with the 5-fold

cross-validation. However, this architecture has the limitation of fails to detect multiple wound instances in the image.

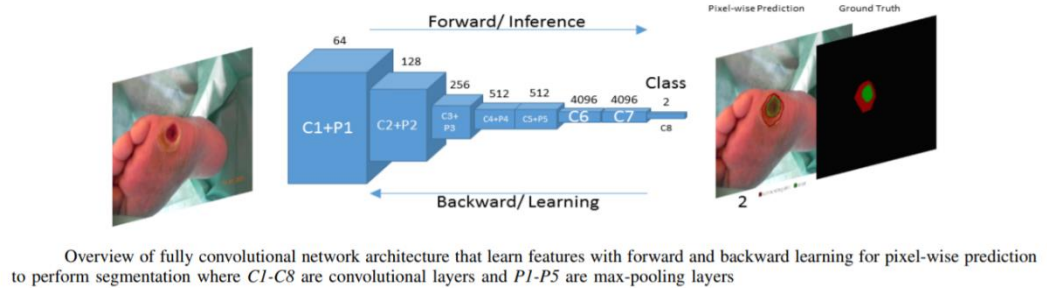


Figure 2.3: Architecture of Goyal et al. Proposed method [33]

Wang et al.[34] proposes a deep convolutional neural network-based approach for automatic wound segmentation. Additionally, they have analyzed the wound surface too. In the automatic wound segmentation task, they propose an end to end styled encoder-decoder deep convolutional neural network architecture as shown in Figure 2.4. The model has achieved 47.3% of mean IoU and 95% of pixel accuracy. The limitation of this architecture is the traditional FCN has no ability to separate multiple objects with a similar class (not instance aware) since this architecture only facilitates to differentiate similar pixels into a single instance.

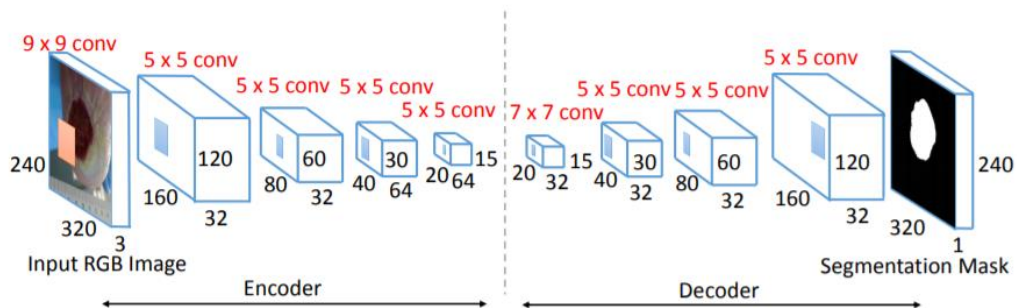


Figure 2.4: Architecture of Wang et al.s' proposed system[34]

Xiaohui et al. [35] propose a deep convolutional neural network architecture -based framework for wound segmentation as shown in Figure 2.5. The proposed framework consists of three modules that facilitate data augmentation, segmentation, and post-processing. Flip, zoom, rotate, and translation was used as the augmentation techniques to avoid the overfitting problem. They have used MobileNet pre-trained architecture with three different mediums of channels as backbones. They have obtained the highest mIoU of 84.60% for MobileNet×0.75-fcn16.

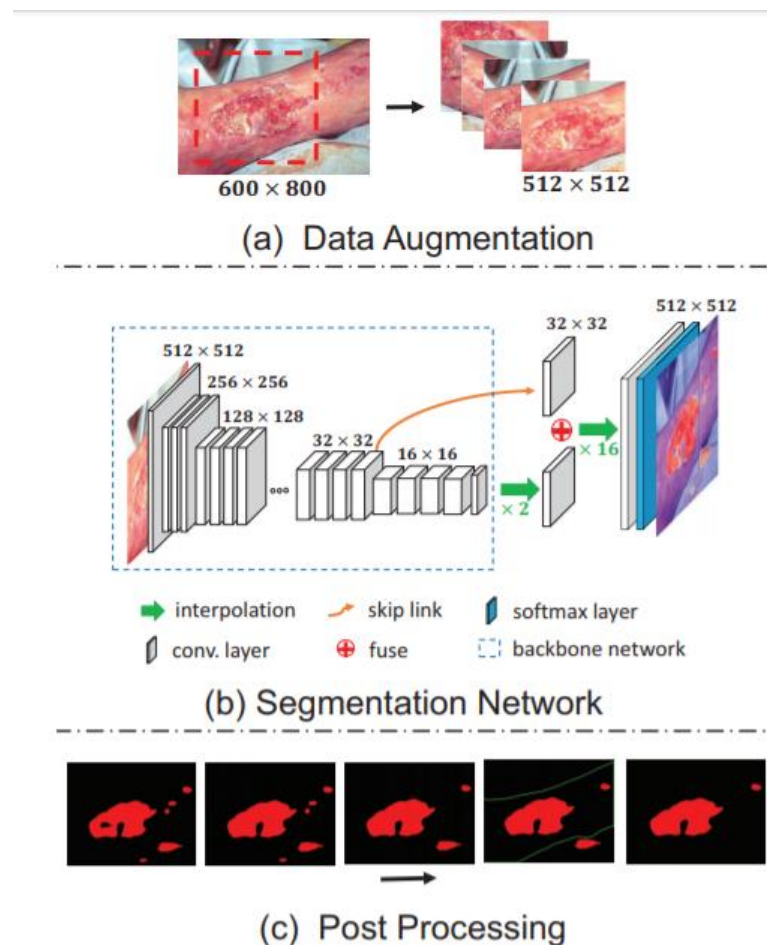


Figure 2.5: Architecture of Xiaohui et al s' research work [35]

In addition to wound segmentation, we found novel approaches to medical image segmentation through deep learning methods.

Ronneberger et al. [36] have published a novel CNN architecture to analyze medical imagery. They introduce a combination of an equal amount of encoder network layers and decoder network layers in their proposed architecture. They use three different tasks namely segmentation of neuronal structures in electron microscopic recordings, HeLa cells on a flat glass recorded by differential interference contrast (DIC) microscopy, a cell segmentation task in light microscopic images for testing the proposed architecture. This model has achieved the mean average precision over 77.5%. Brosch et al. [37] propose an approach for white matter lesions segmentation in brain MRI based on U-Net like architecture. Recurrent Neural Network (RNN) have used in several medical image segmentation tasks. Yuanpu et al. [38] have proposed spatial clockwork RNN based architecture to segment Muscle Perimysium. This research has evaluated using 348 H&E stained skeletal muscle microscopy imagery and they have obtained 0.842 of average IoU.

Deep neural network-based architecture in order to segment the brain tumors has proposed by Havaei et al. [39]. They have used a cascade CNN architecture that exploits both local features and more global contextual features. Furthermore, this architecture consists of a final layer that is a convolutional implementation of a fully connected layer that allows a 40 fold speed up. They have used a two-phase training procedure which allows tackling difficulties related to the imbalance of tumor labels. They have used the MRI image dataset from the BRATS-2012 challenge and they could achieve over 0.80 Dice Similarity Coefficient.

Whitehead et.al. [40] have proposed a deep learning-based architecture which uses a series of four pixel-wise segmentation network to segment the spinal cord. Each network in the series segments MR images at different scale and the results from one network are fed into the next network in the series. Each network uses both the original image and the results from the previous network to produce better segmentation

results. They have used the UCLA Radiology Pictures and Communication (PACS) database for their experiments and they could achieve a Dice score of 0.865.

2.1.5 Image Classification Methods

Goyal et al. [41] propose a deep convolutional neural network for the Diabetic Foot Ulcers classification. However, this automated system limits to binary classification to classify as abnormal skin (skin with diabetic ulcers) and normal skin (healthy skin). The system has been implemented with its own custom convolution network architecture called DFUNet and the model achieved a 0.961 AUC score. The initialization layer inspired by GoogleLeNet, parallel convolution layers to discriminate the DFU more effectively than the previous network layer and both fully connected layers and a softmax-based output classifier are the main three section which appears in this architecture.

There is no research work available in the current literature for Diabetic Foot Ulcer (DFU) classification into severity stages. But we found different deep learning-based architectures that have used in medical image classification tasks and few of them are described below.

Dorj et.al. [42] propose a deep learning-based architecture to classify skin cancers through imagery. In their architecture, they have used the combination of a pre-trained AlexNet convolutional neural network and an ECOC SVM classifier. They have extracted the features of imagery from the AlexNet and have fed into ECOC SVM to do the classification. They have done their experiments using a 3753 image dataset and they could achieve over 91.8% accuracy.

Thambavita et.al. [43] propose an approach to the classification of the anomalies in the gastrointestinal tract based on deep convolutional neural networks. Their proposed architecture is an ensemble of pre-trained ResNet-152 and Densenet-101. They have used four thousand imagery dataset which belongs to eight classes and they could achieve 95.08% accuracy.

Mateen et al. [44] propose an approach to the classification of fundus images using a deep learning-based architecture. In their architecture they have used the Gaussian mixture model for region segmentation, VGG 19 pre-trained architecture as the high dimensional feature extractor, PCA and SVD as feature reduction techniques, and softmax layer as fundus image classifier. They have used the KAGGLE fundus image dataset which contains 35126 fundus images in their experiments and has achieved the classification accuracy over 92.21%.

Zhane et al. [45] have proposed a deep convolutional neural network-based solution to classify gastric precancerous disease namely polyp, erosion and ulcer. They have introduced fire modules from SqueezeNet to reduce the model size and parameters while improving the speed. They have proposed iterative reinforced learning methods in order to maintain the classification accuracy with fewer parameters. Initially, they have trained their model from scratch and then they have applied iterative reinforced learning for fine-tune the hyperparameters. After then they have considered the modified model as a pre-trained model for the next training phase. In their experiments, they have used 1331 imagery and they have obtained 88.90% of classification accuracy.

Nawaz et al. [46] propose a deep CNN based approach for breast cancer classification. Their classification consists of two stages including malignant and benign classification and subtypes classification for both cases. Subtypes of malignant breast tumors are Lobular carcinoma (LC), Ductal carcinoma (DC), Papillary carcinoma (PC) and Mucinous carcinoma (MC) and the subtypes of benign tumors are Fibroadenoma (F), Adenosis (A), Tubular adenoma (TA) and Phyllodes tumor (PT). In order to do the classification, they have used DenseNet based CNN architecture with transfer learning. For their experiments, they have used a 7909 image dataset and they have achieved 95.4% accuracy in the multi-class breast cancer classification task.

2.2 Convolutional Neural Network

This is an architectural wise improvement of an artificial neural network. Most frequently it has been used for image analysis and has been exposed to a good

performance in classification and segmentation tasks. Recently, medical image analysis groups use CNNs and other deep learning algorithms in a wide range of applications, and remarkable outcomes are advancing persistently. Basically, a CNN can be separated into two parts namely feature extraction and the classification as shown in Figure 2.6.

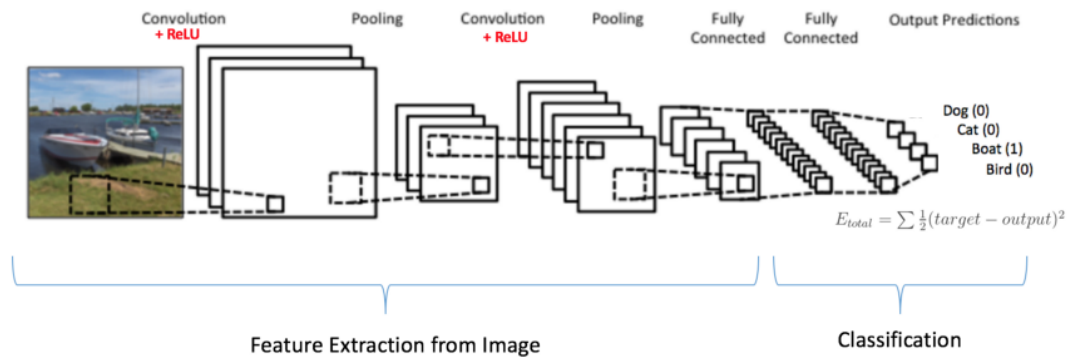


Figure 2.6: Traditional CNN Architecture

2.2.1 Feed-Forward Neural Network Vs Convolutional Neural Network

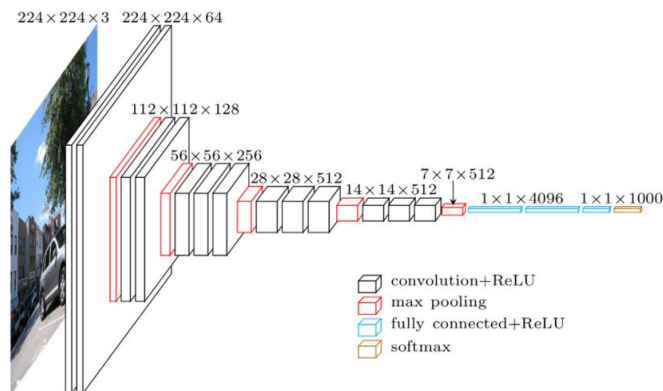


Figure 2.7: Down Sampling

Assume that there is an RGB image with a size of 224 x 224. In this case, the input layer of the neural network will require $224 \times 224 \times 3 = 100,352$ neurons. This is computationally ineffective. Therefore the input tensor dimension can be reduced to 1 x 1 x 1000 (this is called downsampling) by applying the convolutional operation.

Therefore, 1000 neurons will require in the input layer of the feed-forward neural network.

2.2.2 Convolution Operation

First, overlap the kernel to the input map and do the element-wise multiplication and add the result. Figure 2.8 shows the convolution operation in between a 6x6 input map and 3 x 3 kernel and it outputs a 4 x 4 feature map. Figure 2.9 shows the next convolution operation by moving the kernel according to the value of stride. Similarly, the need to perform the convolution operation for the entire input map.

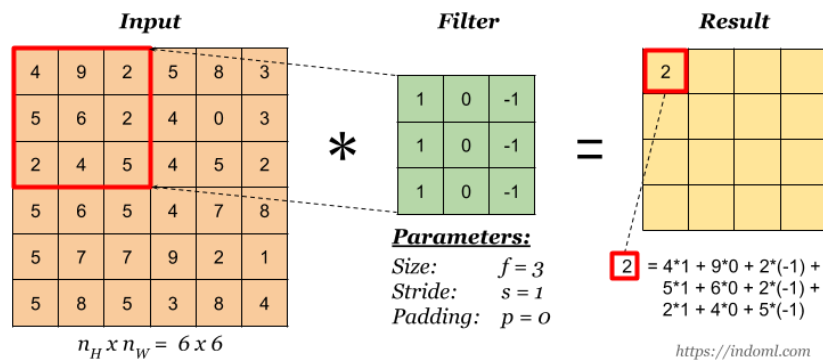


Figure 2.8: Initial Convolutional Operation

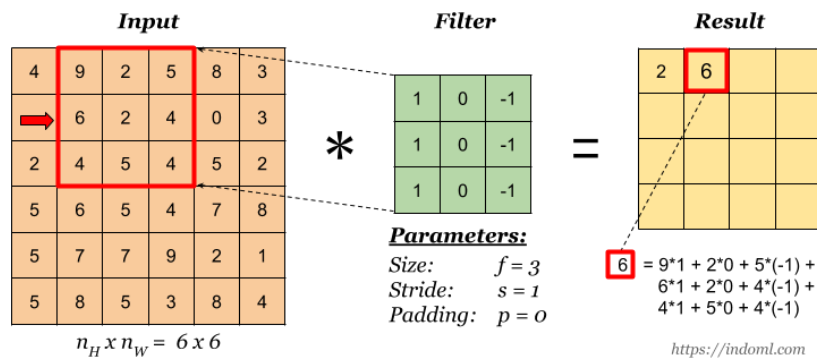


Figure 2.9: Convolution operation after moving by a stride

2.2.3 Input Layer

This layer consists of data of the image. Generally, RGB image data is denoted by a three-dimensional matrix since it has three color channels. Therefore it needs to be reshaped into a vector. If the dimension of the image is $100 \times 100 \times 3$ then it will be converted into 30000×1 and if the batch size is n , the size of the input matrix will be $(30000, n)$.

2.2.4 Convolution Layer

The role of this can be defined as a feature extractor since it extracts the features from the image. In this process, a kernel slides over the image and performs the convolution operation (calculate the dot product between the kernel and the respective field which has the same dimension as the kernel) at each location. The size of the kernel and the stride are hyperparameters and can be predefined and can be different for each convolution layer. The output from the convolutional operation will be feed into an activation layer because without the activation function the whole network would be a series of linear operations, which could be replaced by a single linear operation. In the real-world, most tasks are non-linear so a linear algorithm will not be sufficient to solve such tasks in a satisfactory manner. Most commonly used activation function is ReLU activation which behaved according to the $f(x) = \text{MAX}(\text{zero}, \text{input})$ function. In simply this function converts the negative input values into zeros. This function gives a solution to vanishing gradient problem which has to face while training the network using gradient-based algorithms.

2.2.5 Dropout Layer

The dropout layer is very useful when adhering to the overfitting problem even though this layer is not a part of traditional CNN architecture. This layer forces the network to be redundant by dropping out a random set of activations in that layer. It means that the network should have the ability for accurate classification even though some activations have been dropout so through this it proves that the network is not too fitted on training data.

2.2.6 Pooling Layer

Reduce the spatial volume of the feature map which is an output of the convolution layer is the major purpose of applying the pooling layer. This layer helps to decrease the computational power required to process the data through dimensionality reduction and helps to effectively train the model by extracting the dominant features which are rational and positional invariant. There are four pooling methods namely max-pooling, min-pooling, average pooling, and global average pooling. Max pooling considers the max value from the image portion covered by the kernel region. Min pooling considers the min value from the image portion covered by the kernel region. Average pooling considers the average value from the image portion covered by the kernel region. A visualization of max, min, and average pooling is as in Figure 2.10. Until the satisfactory level of detail, extraction has been achieved these operations are cascaded in layers.

The most commonly used pooling method is Max pooling because it performs as a Noise Suppressant. It has the ability to discard the noisy activations and performs de-noising along with dimensionality reduction.



Figure 2.10: Visualisation of min, max, and average pooling

Global Average Pooling: Global Average Pooling (GAP) is a pooling method that averages all the features within a feature map. Moreover, a given feature map with dimension $h \times w \times d$ reduced into $1 \times 1 \times d$ through GAP. The GAP layer helps to

minimize overfitting problem by reducing the number of learnable parameters in the model. And also used to reduce the spatial dimensions of an n-dimensional tensor. Figure 2.11 shows an example of a feature map with $7 \times 7 \times 1024$ reduces into $1 \times 1 \times 1024$ through GAP.

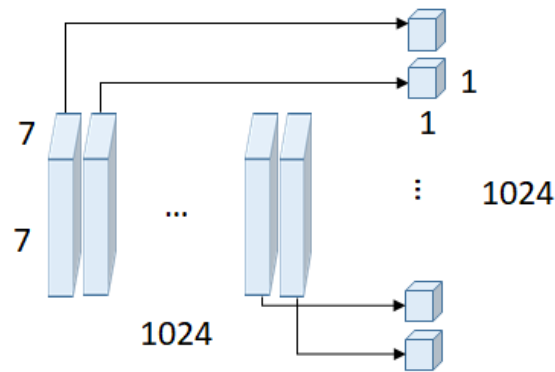


Figure 2.11: Global Average Pooling

2.2.7 Fully Connected Layer

This layer is a traditional neural network that takes the output from the previous layer (Conv layer /pooling layer) as the input and learns non-linear combinations of the high-level features. This layer facilitates to identify which features are present and how they are oriented in the image by using the local features that have been extracted from the first part of the network.

2.2.8 Softmax Layer

Outputs from the FC layer are sent through the softmax layer which is featured by the softmax function. Softmax function squeezes the input vector of real-valued scores into a vector of values between zero and one that sums to one.

2.3 Transfer Learning

Transfer learning is learning of new task using the knowledge which has been obtained from a different task. This is a better solution when there is a lack of training data to

reconstruct models. Even though the source and the target tasks focus on different topics or feature spaces between domain data are different, boosting the performance of the target task [47]. In most cases, the existence of data is poor in medical image processing. Therefore transfer learning is a viable way when CNN employment medical image classification with the help of enough annotated imagery. In [48] have analyzed the mechanisms of deep transfer learning for medical imagery. In [49] has described two methods widely used for transfer learning in deep networks.

Feature extraction from a pre-trained network: The last two fully connected layers of the network are swapped with our own uninitialized layers. The weights of the network are not affected by back-propagating the gradients because they are frozen.

Fine-tuning of a pre-trained network: Usually comparable to the previously mentioned approach but we utilize a little learning rate to fine-tune the weights for the task rather than freezing.

3. DATASETS

3.1 Diabetic Wound Dataset

The dataset consists of imagery that belongs to six classes as mentioned in Table 1.1. The dataset classification was done according to the Wagner score with the help of well-trained clinicians. The entire dataset has 2400 imagery. The dataset contains high-resolution images and captured in different conditions such as lighting, colors, different orientation, etc. The dataset is divided into training and testing where the training set represents 80% and the test set represents 20% from the whole dataset. We use the VGG Image Annotator (VIA) tool in order to annotate the wound images. We export the mapping between image id and the coordinates of the annotated points into a JSON file.

3.1.1 Ground Truth Mask Generation

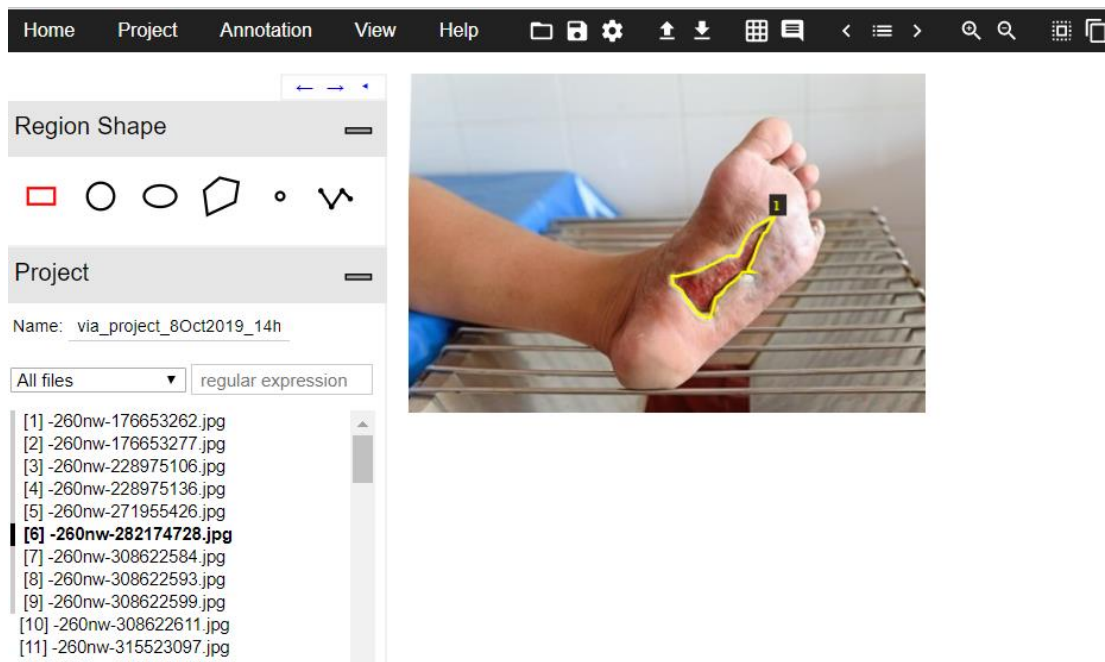


Figure 3.1: User Interface of VGG Annotator Tool

We use the VIA-VGG image annotator tool [50] in order to create ground truth masks of ulcer images. An example of this utility is shown in Figure 3.1. This annotator tool is a simple one and it is a single HTML file that enables download and opens in a

browser. The clinician draws the bounding polygons around regions of wounds and the wound regions are labeled with 1 and the background region got 0 by default. The metadata of the annotations was stored in a JSON file by the tool itself.

We use another two different medical datasets additionally to the diabetic wound dataset, to validate our results of the segmentation task. Section 3.2 and 3.3 describe the details about these datasets.

3.2 CVC-Colon Dataset

This dataset [51] has introduced with the purpose of polyps detection. This has the colonoscopy video sequences along with their annotations. There are 300 polyps images and their binary masks in the dataset.

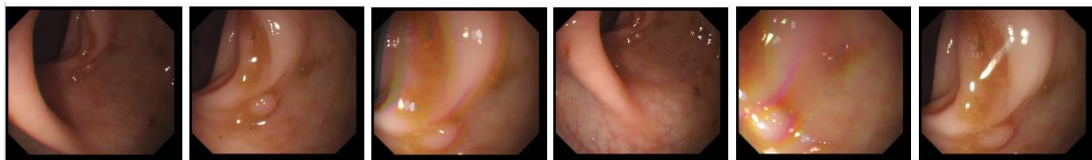


Figure 3.2: Samples of Polyps dataset

3.3 Chest X-ray Dataset

The size of the dataset is 662 chest x-ray images that have collected from Shenzhen Hospital. The manually segmented binary mask images for this dataset were collected from Lung Masks for Shenzhen Hospital Chest X-ray Set given by Kaggle[52]

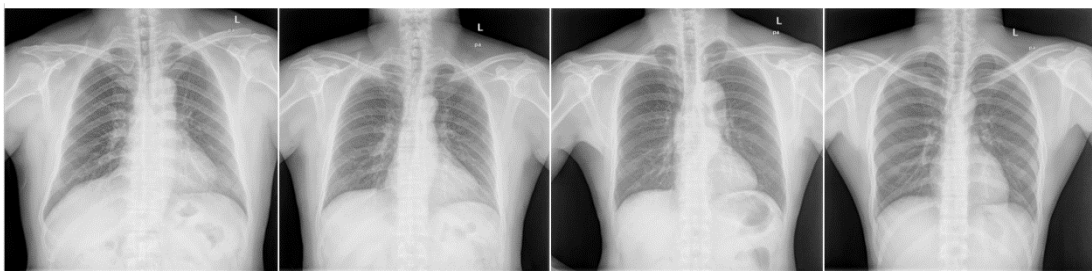


Figure 3.3: Samples of Chest X-ray dataset

4. METHODOLOGY

The purpose of this chapter is to show the research methodology of the thesis with the strategy of the research and the empirical techniques applied. First, we are going to analyze the proposed architectures for diabetic wound segmentation tasks and diabetic wound classification tasks. Moreover, this chapter will describe the approach of model selection.

4.1 Task 1: DFU segmentation

4.1.1 Preprocessing

We augment the images in our training dataset as a solution to the overfitting problem. Two types of data augmentation techniques called on-line data augmentation and off-line data augmentation. In this task, the off-line data augmentation technique was used. Therefore we augment and annotate the dataset before the training process. Following data augmentation techniques were applied in our experiments.

- Gaussian blur with a standard deviation of 2.5.
- The horizontal image flips.
- Luminosity scaling (within [0.8, 1.5]).
- Image rotations.

The above transformations do not alternate the basic characteristics of the neuropathic ulcer imagery, subsequently, all the transformations valid for performing the augmentation.

4.1.2 Model Architecture

The mask-RCNN [6] model was introduced for instance based segmentation of objects. This model is an extension of Faster-RCNN architecture. Fast-RCNN is a combination of CNN backbone networks and an RPN (Region Proposal Network). Fast R-CNN provides two major outputs for each object whereas the anchor box and the class label of the segmented object. Mask R-CNN outputs an object mask that uses

to do the object segmentation more precisely additionally to the outputs of Faster R-CNN.

Initially, this scans the image and fed into a backbone model to extract the feature maps. These feature maps fed into an RPN which has the ability to predict if there is an object or not in the selected region and it generates the bounding boxes for the region of interest. Through the ROI align technique, resize the generated candidate bounding boxes as the same size of candidates. Then this fixed vector is parallely fed into fully connected layers and a regression layer in order to predict the bounding box and the class label. And also this vector is fed into the mask branch and it outputs the segmentation mask for each region which consists of an object. The multi-task loss function of the model is as follows,

$$L = L_{cls} + L_{bbox} + L_{mask}$$

$$L_{cls} = \text{rpn_class_loss} + \text{mrcnn_class_loss}$$

$$L_{bbox} = \text{rpn_bbox_loss} + \text{mrcnn_bbox_loss}$$

$$L_{mask} = \text{mrcnn_mask_loss}$$

rpn_class_loss: this describes the quality of separation between the objects and background by the Regional Proposal Network.

rpn_bbox_loss: this describes the quality of object localization by the Regional proposal network.

mrcnn_class_loss: this describes the level of recognition of each class object by the Mask-RCNN.

mrcnn_mask_loss: this describes the level of object segmentation by Mask R-CNN.

mrcnn_bbox_loss: this describes the level of object localization by Mask R-CNN.

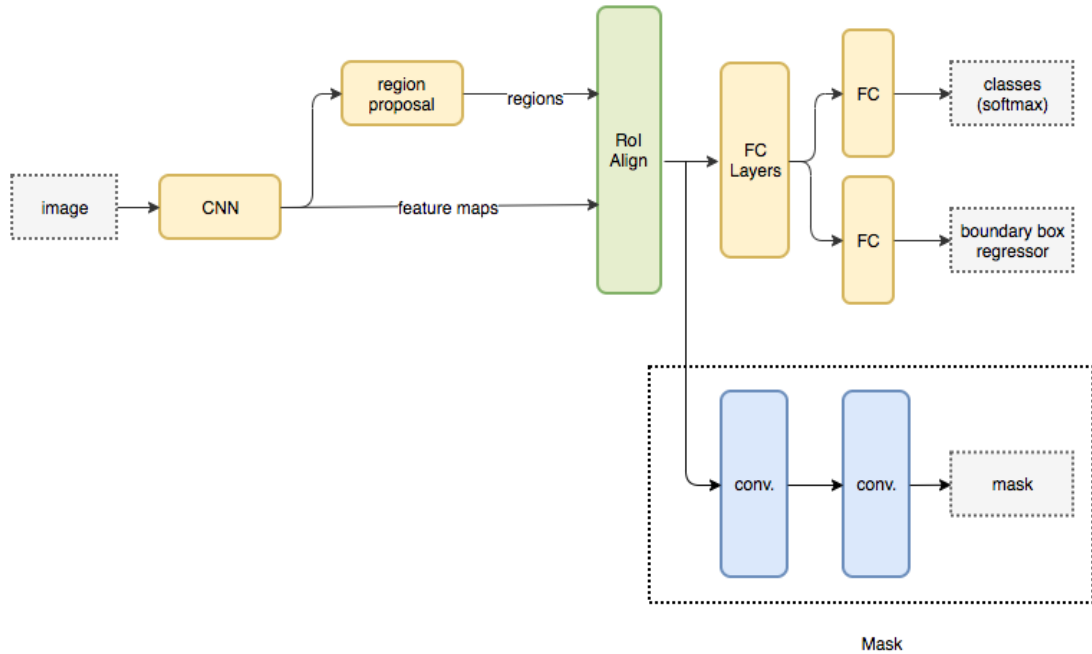


Figure 4.1: Mask R-CNN Architecture [6].

4.1.3 Model Training

We use ResNet-50 and ResNet-101 backbone networks for our experiments of model training. Table 4.1 describes the optimal hyper-parameter configuration that we obtained from different experiments in the training phase. For model initialization, we used the pre-trained weights using the MSCOCO dataset [53]. We use three stages to train the model. The head layers (the RPN, classifier and mask heads) of the network were trained during the first 30 epochs and then the upper layers (4+ layers) of the network were trained in the next 50 epochs. Finally, the entire network architecture was fully trained during the last 20 epochs. In total, we used 100 epochs for model training and the optimizer of the model was Stochastic Gradient Descent (SGD).

4.1.4 Hyperparameter Tuning

Table 4.1: Hyperparameter Table

Number of Classes	2
Backbone	ResNet 101
Image per GPU	2
RPN anchor scale	32, 64, 128, 256, 512
Train ROIs per image	128
Mask Shape	28 x 28
Anchors per image	256
Learning rate	0.001
Learning Momentum	0.9
Weight Decay	0.0001
Batch Size	2

We performed the segmentation between ulcer pixels and the background pixels so that the value of the Number of Classes parameter was 2. We used ResNet-101 and ResNet-50 as the backbone architectures and according to our comparisons, in terms of model performance, ResNet-101 was the best-performed backbone model. Learning momentum was 0.9 for the SGD optimizer and the L2 regularization (weight decay) parameter was 0.0001. We do our experiments within the range of 0.01 to 0.0001 of the learning rate. The optimal value for the learning rate was 0.001 and there was no noticeable improvement for other values. According to our memory of the machine, the rest of the hyperparameters were chosen.

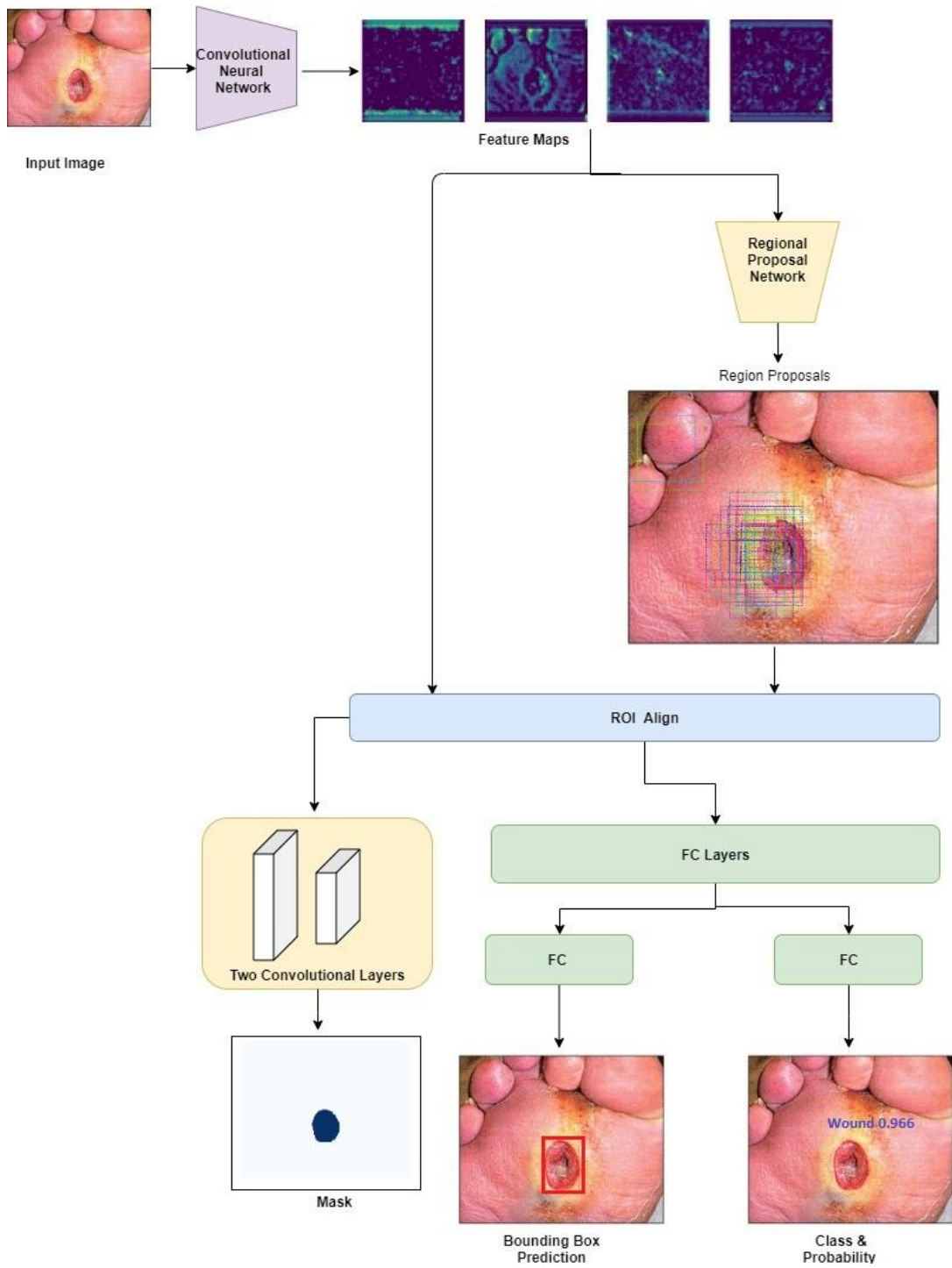


Figure 4.2: Workflow diagram of Mask R-CNN

4.1.5 Reasons for choosing Mask R-CNN architecture

CNN has the strength of image classification and object detection using bounding boxes. Therefore, we are able to use CNN in image segmentation tasks. It has limitations because it uses bounding boxes for object detection but for only one object at a time and it doesn't work well when there are multiple objects in the visual frame due to interference. So we need a solution to identify multiple objects in one frame and draw the related bounding boxes separately around the objects. As a solution, we found region-based CNN (R-CNN) from the literature and it has the ability to the detection of multiple objects in one frame.

R-CNN (Ross Girshick et al. 2000)

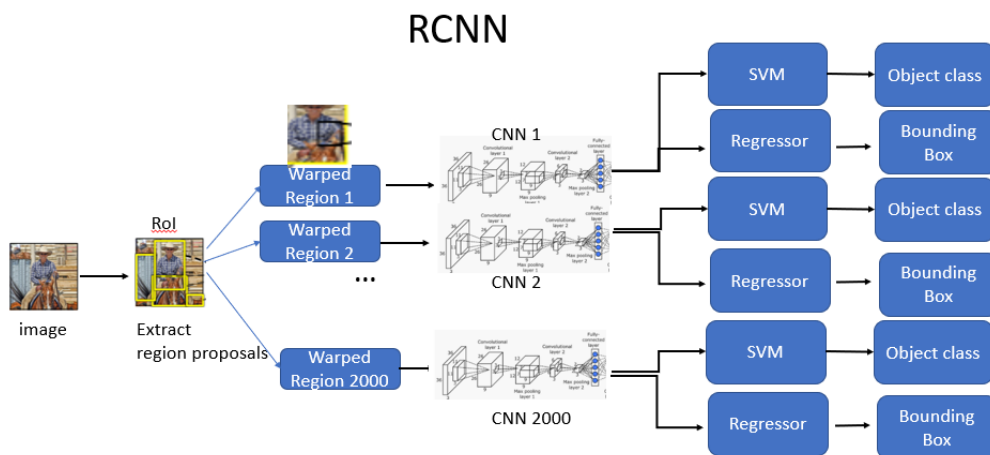


Figure 4.3: R-CNN Architecture

The selective search algorithm has used in order to extract 2000 ROIs for each image in R-CNN. All 2000 of ROIs wrap into a square. Then each ROI is sent separately through the CNN for feature extraction and it outputs a 4096-dimensional feature vector. This feature vector fed into an SVM to identify the class of the object and fed into a regressor to predict bounding box.

This algorithm has architectural issues because it requires high processing time to train the model since it has to classify 2000 ROIs per image. And also the selective search

is a fixed algorithm and there is no learning part in the ROI generation step so it will lead to the generation of bad candidate ROIs. This architecture consists of three models namely CNN, linear SVM, and regression model so this leads the training process complex.

Fast R-CNN

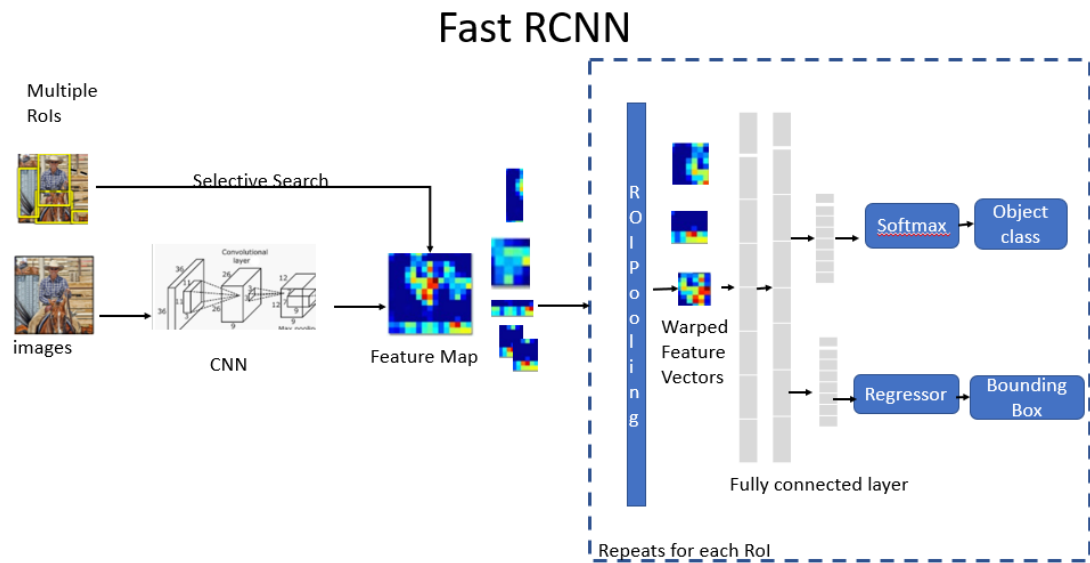


Figure 4.4: Fast R-CNN Architecture

In order to overcome the issues in R-CNN architecture, the Fast R-CNN is proposed by the same authors of R-CNN. Figure 4.4 indicates the architecture diagram of Fast R-CNN. This architecture is an extended version of R-CNN. Instead of feeding 2000 ROIs per image to CNN, it takes the entire image at once for feature extraction by CNN. Using this feature map and manually generated ROIs from the selective search algorithm will be wrapped into squares through the ROI pooling. This reshapes feature vectors fed into the fully connected layers and it uses the bounding box regressor and softmax layer to predict the bounding box and class label. This architecture is more efficient than the R-CNN because it uses a single deep CNN for feature extraction of the entire image at once, it uses a softmax layer that outperforms SVM for object classification, and it uses multi-task loss function for training deep CNN which increase the accuracy.

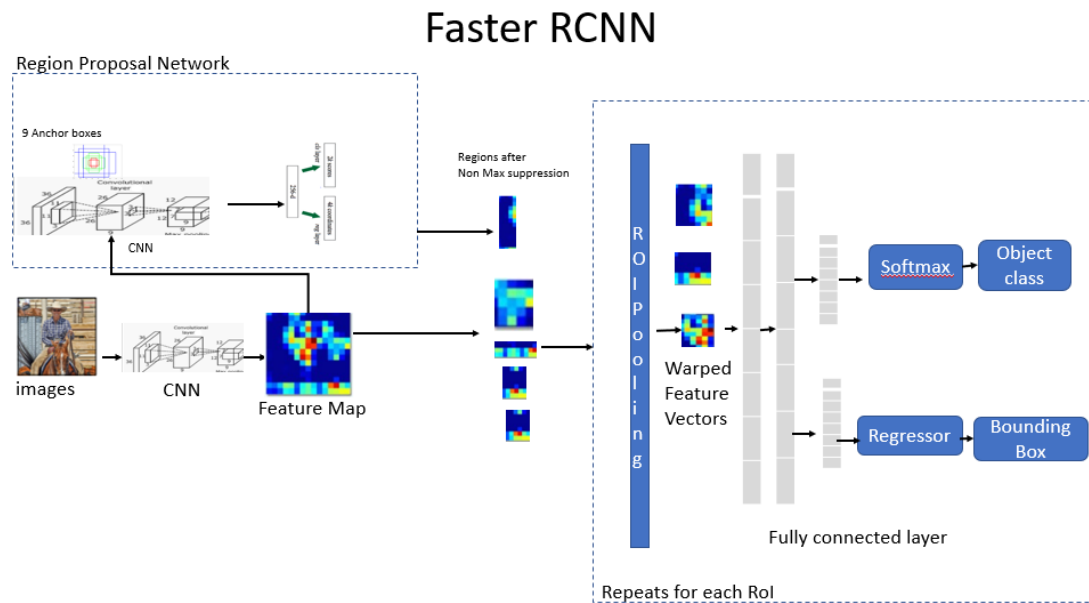


Figure 4.5: Architecture of Faster R-CNN

This model consists of an improved version of Fast R-CNN architecture and it addresses some issues that exist in Fast R-CNN. This replaces the selective search algorithm from the regional proposal network to generate the ROIs dynamically. Faster R-CNN model gets an image as the input and is feed into the CNN to generate feature maps. The separate regional proposal network is used to generate bounding boxes of object and objectness scores. ROI pooling layer is used to reshape the ROIs and wrap them into a fixed vector size. Then this vector is parallely fed into a regression layer and fully connected layers In order to predict the class label and bounding box.

We explore the image segmentation concept from a simple CNN to Mask R-CNN. We identified that RCNN is very slow and expensive and Fast R-CNN uses a selective search mechanism for generating region proposals. Even though the Faster R-CNN overcome all those issues, it supports only for the semantic segmentation, but not the instance segmentation. But in our task, we need to segment multiple wounds separately in one image and so we need instance segmentation to perform it. Therefore

considering these issues and the limitations, we identify the Mask R-CNN is the best-suited model for DFU segmentation

4.2 Task 2: DFU Classification

4.2.1 Preprocessing

Our dataset contains the images in different resolutions. Therefore we downsample the imagery into 224 x 224 pixels because it would be feasible for any CNN to converge in a reasonable time.

4.2.2 Addressing the class imbalance problem

Class imbalance problems can be addressed through two approaches; using class balanced training sets or using cost-sensitive learning.

The training dataset that we use for our experiments has unbalanced class distributions as shown in the Table 4.2. This makes an additional overhead to our model. In order to prevent this issue, we use a cost-sensitive learning approach in the training process. We did not apply the data augmentation technique in order to balance the class distributions since we used the stratified cross-validation technique for the model training.

In this method, we initially adjust weights as inversely proportional to class frequencies and then passed into the fit function of the models during the training stage which has done using stratified cross-validation. Therefore each fold contains approximately the same proportion of observations as in the training dataset.

Table 4.2 : Class Distribution of Training Dataset

Class (Severity Stage)	Number of Images
Grade 0	303
Grade 1	353

Grade 2	287
Grade 3	294
Grade 4	325
Grade 5	304

4.2.3 Approaches

4.2.3.1 DFU Image Classification through Fine-Tuning CNN Models

In this approach, we perform the classification through fine-tuning different pre-trained CNN models. Where CNN models are DenseNet[8], ResNet[54], Xception[55], InceptionV3[56], InceptionResNetV2[57] and VGG[58]. The steps of the approach are as below. First, we preprocess our dataset. Then we developed a pre-trained CNN up to the last fully connected layer. We replace the last fully connected layer with a softmax layer which contains 6 neurons to do our DFU classification. We initialize the weights of the input layer to the last pooling layer with ImageNet weights and weights of the custom softMax layer using the He normal initialization. We fine-tune the network up to 400 epochs with a Stochastic Gradient Descent (SGD) optimizer with 0.9 momentum using a mini-batch size of 32. We use 0.001 learning rate, 0.0005 L2 regularize parameter. In this approach, we achieved the uppermost accuracy of 71.34% for DenseNet 201 architecture.

4.2.3.2 DFU Classification through Feature Extraction

In this approach, we use each aforementioned pre-trained CNNs as a feature extractor and we applied the Global Average Pooling (GAP) layer to reduce the size of the extracted feature vector. Totally 1920 features were extracted from DenseNet-201 CNN after applying the GAP layer. Then we normalize and apply Singular Value Decomposition (SVD) on this feature vector to reduce the dimensionality. In order to

select the optimal number of features (222 features), we consider the 95% variance threshold for truncated SVD by plotting the cumulative sum of the singular values against the principle components. These extracted features were used to train different classifiers namely artificial neural network (ANN), Random Forest, Support Vector Machine (SVM). We use stratified 5-fold cross-validation for model training.

The pre-trained CNN based on DenseNet 201 architecture as the feature extractor with a single hidden layer ANN classifier which consists of 128 neurons with ReLU activation functions gave the best accuracy of 96.22%.

The randomized search strategy was used for the selection of hyperparameters in our ANN model. We consider the training and validation errors that were obtained from stratified 5-fold cross-validation.

In accordance with the theoretical aspects, if the difference between training and the validation error is high and the training error is low, it indicates that the model is failed to generalize new samples. in order to solve the overfitting problem, we use regularization techniques such as L2 regularization, dropout, normalize extracted features and batch-normalization.

Moreover, if the performance of the training dataset on the cost function is very low, we increase the number of epochs to run the gradient descent longer. In this experiment, we use 0.001 of the learning rate, 400 epochs and 0.4 dropout values for all scenarios because varying of the values did not make a noticeable impact on the results. Furthermore, we test with changing the number of hidden layers in the ANN and neurons per layer, but we got the best result for a single hidden layer ANN with 128 neurons with the ReLU activation function. The Overall architecture of our best performing approach is as shown in Figure 4.6.

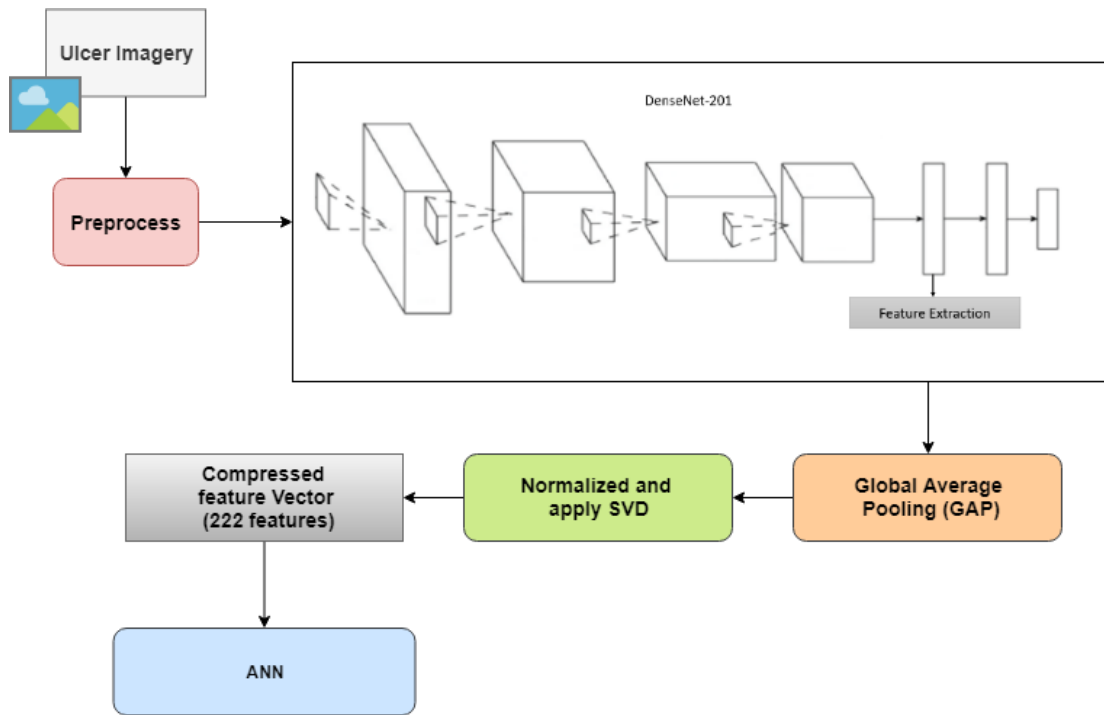


Figure 4.6: DFU Classification Architecture

5. RESULTS & EXPERIMENTAL ANALYSIS

5.1 Results of DFU Segmentation

5.1.1 Evaluation Metrics

The evaluation metrics for the DFU segmentation task, we consider the Average Precision (AP) at different thresholds of intersection over union (IoU). The IoU of a set of the predicted set of object pixels and ground truth object pixels is calculated as:

$$IoU(Target, Prediction) = \frac{(Target \cap Prediction)}{(Target \cup Prediction)}$$

In order to calculate the average precision value, we consider the range of IoU threshold values I.e. ranging from 0.5 to 0.95 with a stage size of 0.05 and at each point. Theoretically, a precision value is determined by false positives (FP), false negatives (FN), and true positives (TP) of comparing the ground truth and the predicted masks at each threshold value t .

The mean average precision of a single image is then calculated as the mean of the above-average precision values at each IoU threshold.

$$mAP = \frac{1}{|threshold|} \sum_t \frac{TP(t)}{TP(t) + FP(t) + FN(t)}$$

TP: The obtained IoU score where IoU value of target and prediction mask pair exceeds some pre-defined threshold value.

FN: we define this when the no predicted mask for the object, but we have a ground truth object mask

FP: we define this when there is a predicted mask from the model but there is no relative ground truth mask for that.

Example: At 0.5 IoU threshold

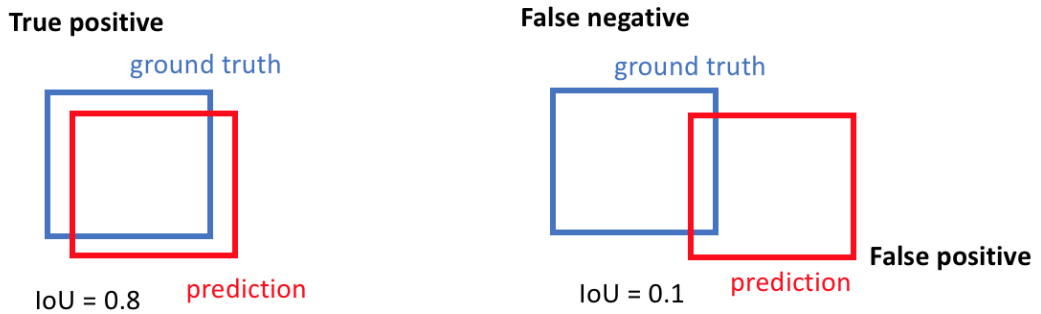


Figure 5.1: Visualization of TP, FP, and FN

5.1.2 Comparison Model

We use the U-Net[59] model as our comparison model. This model consists of deep CNN architecture. Generally, this model is used for small medical datasets and it has shown fairly good performance in medical image segmentation tasks.

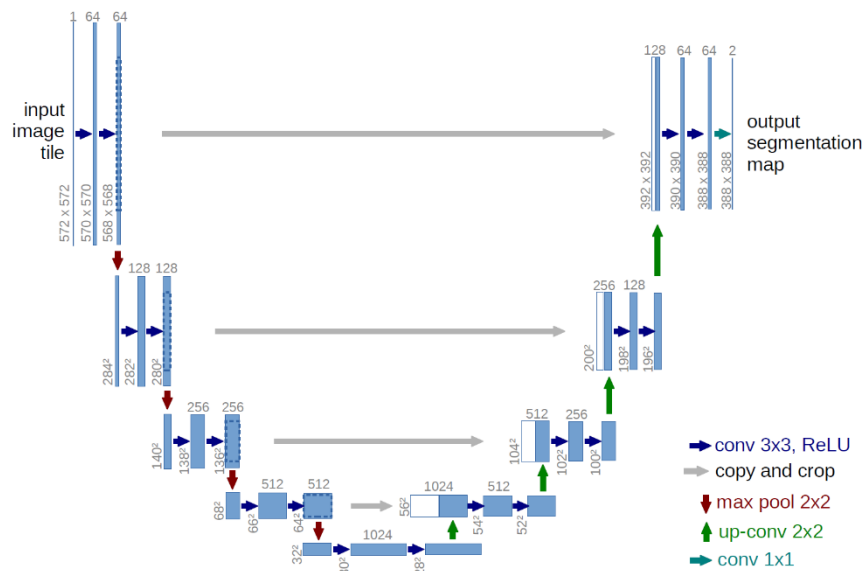


Figure 5.2: U-Net Architecture [59]

Figure 5.2 indicates the architecture of this model. According to the figure, it demonstrates that the network merges a CNN architecture with a deconvolutional architecture. The CNN contains two consecutive convolution layers where each layer consists of 3 x 3 filters followed by a ReLU layer and a 2 x 2 max-pooling layer. This block pattern is repeated multiple times. A 3 x 3 CNN layers followed by a 2 x 2 up convolution layer were used at the bottommost layer. This layer mediates between the CNN and deconvolution network. The deconvolutional network consists of an upsampling operation on feature maps. This network has blocks that contain two 3 x 3 CNN layers and 2 x 2 upsampling layer. The whole network is 23 convolutional layers, where the last layer is utilized to outline segments include vector identified with the number of classes. We train this model up to 100 epochs with Adam optimizer where the learning rate was 0.001. The Xavier initialization scheme was used for weight initialization.

5.1.3 Results on DFU Dataset

We have accomplished 0.1564 mAP for the U-Net model and the average precision at IoU 0.5 was 0.4672 and IoU 0.75 was 0.3041. We achieved 0.4392 mAP with ResNet-50 backbone for the Mask-RCNN model. The average precision at IoU 0.5 was 0.8281 and at IoU 0.75 was 0.5176 respectively. The ResNet-101 backbone architecture demonstrated an mAP of 0.5084 with the Mask R-CNN model. The average precision at IoU 0.5 was 0.8632 and at IoU 0.75 was 0.6157. Table 5.1 shows a summary of the results. The ResNet-101 backbone architecture provided the best results for the dataset.

Figure 5.4 demonstrates the detailed results of boundary detection on few images from our test dataset : (a) Input image (b) Ground truth mask (c) Predicted Wound Mask on ulcer (d) Predicted Binary Mask (e) Predicted contour of ulcer. Figure 5.3 illustrates a few more segmentation results. The red contour indicates the predicted boundary and green contour indicates the ground truth boundary.

Table 5.1: Results of DFU segmentation

Model	mAP@[IoU = 0.5-0.95]	AP@IoU = 0.5	AP@IoU = 0.75
U-Net	0.1564	0.4672	0.3041
Mask-RCNN(Backbone=ResNet-50)	0.4392	0.8281	0.5176
Mask-RCNN(Backbone=ResNet-101)	0.5084	0.8632	0.6157



Figure 5.3: Few results of ulcer boundary segmentation task.

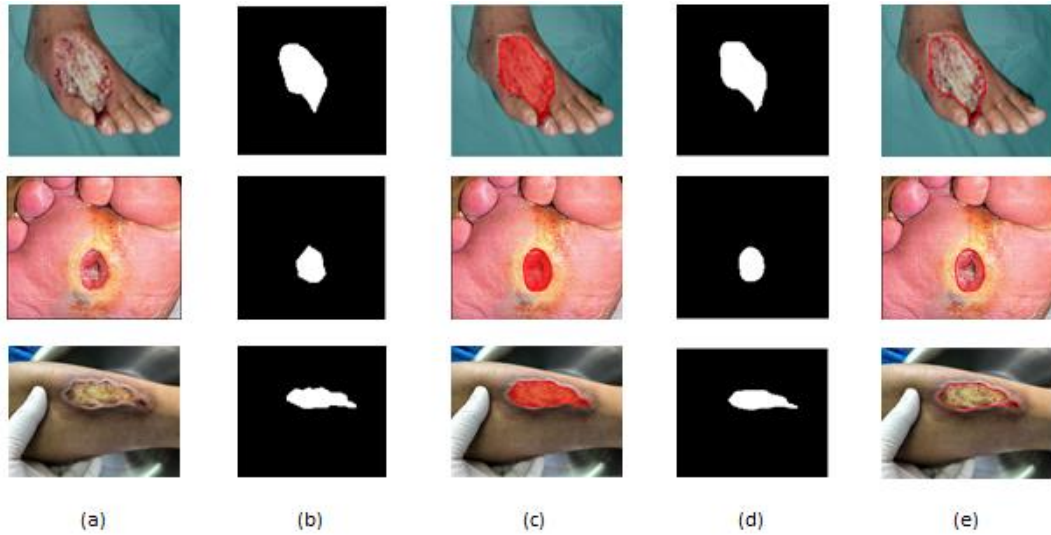


Figure 5.4: Segmentation results for several neuropathic ulcers

5.1.4 Results on Different Medical Datasets

We use two additional medical datasets namely polyps dataset and chest X-ray dataset in order to compare the model performance against the baseline model. We fed each dataset together with their annotations and train the Mask-RCNN model instance separately. Similar configurations were applied for these two experiments as mentioned in Table 4.1. We use the same proportion as mentioned in section 3.1 in order to split these datasets.

Table 5.2: Results of different medical datasets

Dataset	Model	mAP@[IoU = 0.5-0.95]	AP@IoU = 0.5	AP@IoU = 0.75
Chest X-ray	U-Net	0.3021	0.5632	0.4712
	Mask R-CNN	0.6135	0.8715	0.7230
Polyps	U-Net	0.2896	0.5017	0.4321
	Mask R-CNN	0.5031	0.7546	0.6013

Table 5.2 summarizes the obtained results for chest X-ray and polyp datasets. Furthermore, Segmentation results on selected query images for both datasets are visualized in Figure 5.5 and Figure 5.6. Where red contour signifies the predicted boundary and the green contour signifies the ground truth boundary.

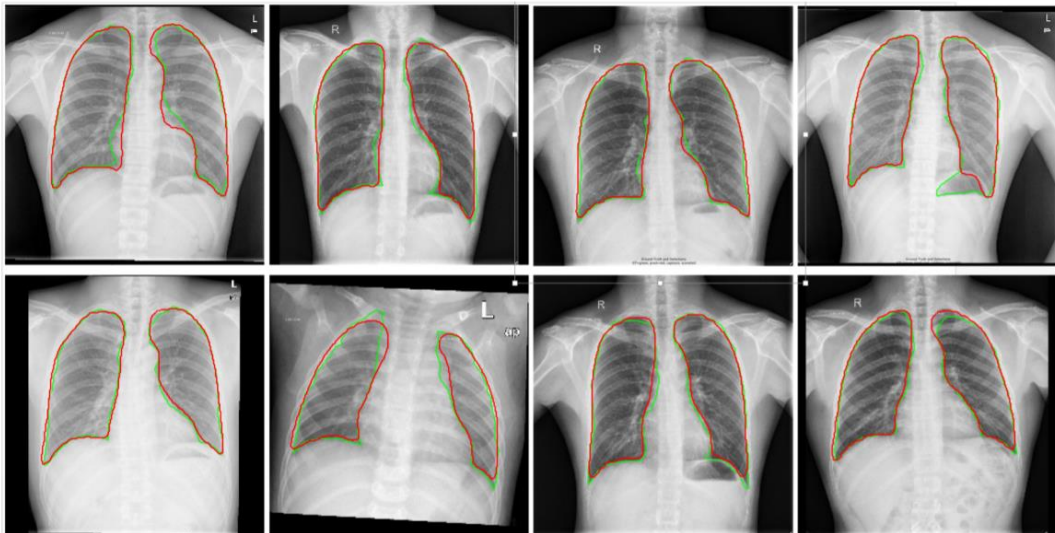


Figure 5.5: Segmentation Results for Chest X-ray dataset with Mask R-CNN

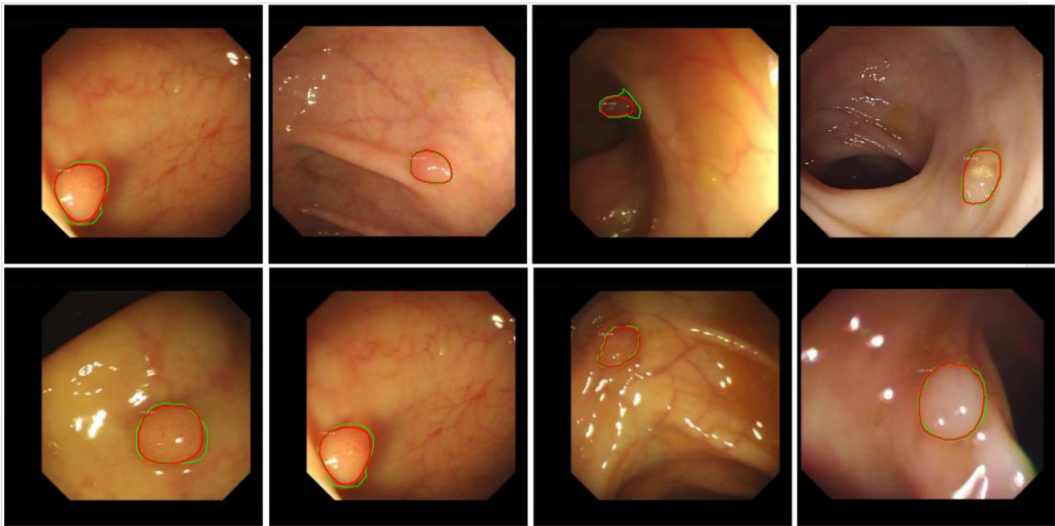


Figure 5.6: Segmentation Results for Polyps dataset with Mask R-CNN

5.2 Results of DFU classification

5.2.1 Evaluation Metrics

In this study, we consider two main metrics for evaluating our models: Accuracy and F1 Score.

1. Accuracy: that measure the proposition of correctly classified samples.
2. $F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$

In order to obtain the best-suited architecture for the DFU classification in terms of accuracy, we perform the experiment with six pre-trained CNNs for the aforementioned (see Section 4.2.3) both approaches. We experiment with three classifiers: ANN, SVM, and random forest in combination with CNN feature extractors. The accuracy and the f1-score for each experiment separately have explained in Table 5.3, Table 5.4, Table 5.5, and Table 5.6.

5.2.2 Results of Feature Extraction through CNN Models for DFU Classification

In the first experiment, we extract the features from CNNs and directly feed them into classifier (see flatten column in tables). As an example in this scenario, we extract $7 \times 7 \times 1920$ features from DenseNet-201 CNN and feed into a classifier. In the second experiment, we apply global average pooling and normalize the extracted features and feed into a classifier (see GAP column in the table). For example, DenseNet-201 extracts $1 \times 1 \times 1920$ features. In the third experiment, we apply GAP to normalize features and next apply truncated SVD, then feed into a classifier (see the results in GAP+SVD column in tables). We use stratified 5-fold cross-validation for all the experiments and also the SGD optimizer gave more accuracy than the Adam optimizer. The summarized results are shown in Table 5.3, Table 5.4, Table 5.5. We achieved the best accuracy for the Densenet-201 feature extractor together with the GAP layer and SVD and the single hidden layer ANN classifier.

Table 5.3: Accuracy and F1-score of CNN feature extractor + ANN

Pretrained CNN	GAP + SVD		GAP		Flatten	
	Acc.	F1	Acc.	F1	Acc.	F1
DenseNet-201	96.22%	0.9610	94.54%	0.9411	52.35%	0.5083
ResNet-18	95.85%	0.9573	91.60%	0.9028	61.56%	0.5078
VGG-16	96.06%	0.9581	90.28%	0.9001	55.36%	0.4703
InceptionV3	95.85%	0.9564	84.64%	0.8144	57.89%	0.4851
Xception	95.58%	0.9559	84.62%	0.8224	61.37%	0.4923
InceptionResnetV2	95.79%	0.9563	83.25%	0.8123	62.15%	0.5310

Table 5.4: Accuracy and F1-score of CNN feature extractor + SVM

Pretrained CNN	GAP + SVD		GAP		Flatten	
	Acc.	F1	Acc.	F1	Acc.	F1
DenseNet-201	64.02%	0.6128	92.94	0.9212	59.43%	0.4915
ResNet-18	62.71%	0.6245	60.31%	0.5963	55.41%	0.5502
VGG-16	58.32%	0.5722	57.32%	0.5639	56.49%	0.5592
InceptionV3	56.96%	0.5502	55.64%	0.5544	54.39%	0.5351
Xception	52.15%	0.5186	51.69%	0.5083	50.20%	0.4984
InceptionResnetV2	48.96%	0.4769	46.21%	0.4596	42.34%	0.4220

Table 5.5: Accuracy and F1-score of CNN feature extractor + Random Forest

Pretrained CNN	GAP + SVD		GAP		Flatten	
	Acc.	F1	Acc.	F1	Acc.	F1
DenseNet-201	69.20%	0.6953	67.25%	0.6698	65.42%	0.6498
ResNet-18	65.69%	0.6483	62.98%	0.6159	61.56%	0.6075
VGG-16	64.35%	0.6325	63.28%	0.6258	60.45%	0.5989
InceptionV3	60.76%	0.5945	58.67%	0.5762	56.39%	0.5548
Xception	58.26%	0.5820	56.32%	0.5596	54.37%	0.5364
InceptionResnetV2	55.45%	0.5486	52.96%	0.5220	50.63%	0.5011

5.2.3 Results of Fine-tuning Pretrained CNN Models for DFU Image Classification

Table 5.6 summarizes the results of the transfer learning approach done with different pre-trained CNNs using ImageNet weights.

Table 5.6: DFU dataset with SGD optimizer results

Pretrained CNN	Accuracy	F1-Score
DenseNet-201	71.34%	0.6520
ResNet-18	69.43%	0.6213
VGG-16	68.80%	0.5997
InceptionV3	61.70%	0.5407
Xception	65.64%	0.5754
InceptionResnetV2	64.58%	0.5772

5.2.4 Comparison Models

In order to measure the performance of the proposed approach with respect to the existing deep feature extraction based methods, we use three deep learning models from the literature and we evaluate them using our DFU dataset. This section describes the models that we use for the comparisons and Table 5.7 summarizes the obtained results.

Method 1: AlexNet Feature Extractor + multiclass linear SVM

Doaa et al.[60] proposed a feature extraction based CNN architecture to classify skin lesions. They proposed a pre-trained CNN AlexNet [60] architecture as the feature extractor and a multi-class linear Support Vector Machine (SVM) has trained on extracted features from CNN as the classifier. We used the same architecture as proposed in [60] with our DFU dataset and we achieved the best classification

accuracy with the linear kernel when the soft-margin parameter (c) = 0.6. We used the same procedure with stratified cross-validation to address class imbalance problem while training this model.

Method 2: AlexNet Feature Extractor + LDA + SVM

Romany F et.al.[61] proposed a feature extraction based architecture to classify diabetic retinopathy. In their, approach they verified that AlexNet DNN-based feature extractor with LDA (Linear Discriminant Analysis) feature selection method and SVM classifier exhibits the best performance in the DR classification task. We trained the same architecture on our DFU dataset. We obtained the best classification accuracy with polynomial kernel-based SVM. We used the same procedure with stratified cross-validation to address class imbalance problem while training this model.

Method 3: VGG-19 feature extractor + SVD + Softmax

Mateen et al. [62] proposed a feature extraction based architecture in combination with VGG-19 feature extractor, SVD (singular value decomposition) feature extractor and softmax classifier for diabetic retinopathy classification. We implemented the same architecture and fine-tune in our dataset. We achieved the best accuracy for Adam optimizer with a 0.001 learning rate in 200 epochs. Weights were initialized by using the Xavier initialization method. We used the same procedure with stratified cross-validation to address class imbalance problem while training this model.

5.2.5 Comparison Results on Proposed Approach and Comparison Models

Table 5.7 summarizes the results obtained against comparison models. The table verifies that compared to the state-of-the-art approaches, our proposed approach demonstrates a significant performance.

Moreover, our proposed architecture consists of 37.47%, 25.59% and 30.81% improvements in terms of accuracy compared to the approaches proposed in Doaa et al.[60], Romany F et.al.[61], and Mateen et al. [62] respectively.

Table 5.7: Proposed approach with comparison models

Model	Accuracy	F1-Score
Densenet-201 + GAP + SVD + ANN (Proposed Approach)	96.22%	0.9610
AlexNet + SVM (Method 1)	58.75%	0.5749
AlexNet + LDA + SVM (Method 2)	70.63%	0.6956
VGG-19 + SVD + Softmax (Method 3)	65.41%	0.6502

6. CONCLUSIONS

6.1 Contribution

A diabetic foot ulcer is a common complication of diabetes. According to the literature, deep convolutional neural networks have been proven great performance in analyzing medical imagery. In order to explore this purpose further, we investigate the ability of CNNs to classify the diabetic foot ulcers into severity stages. DFU classification task introduces a deep CNN based approach for fully automated analysis of DFU imagery. Automate the severity classification of DFU becomes challenging mainly due to the constraints of data availability and the quality of the imagery. Most of the misclassified imagery was with bad lighting conditions. Appropriate stability between the accuracy and the speed of the model is required since the most accurate approaches might not perform well with adequate speed and vice versa.

Here, we apply CNNs based transfer learning for a small dataset. We try out different transfer learning methods for the DFU classification task. In the first attempt, we implant the parameters of different pre-trained CNN models into the domain transferred CNNs which are trained using ImageNet. After doing the weight initialization step we fine-tune transferred CNNs by using DFU images. This approach gave low accuracy and required more computational time. Therefore we discarded this method.

Then we navigate to the feature extraction approach, there we consider pre-trained CNN models as feature extractors, GAP layer and SVD were applied on the DFU dataset and did the feature extraction while preserving the accuracy. Then the feature vectors fed into a classifier. In order to choose the best classifier, we train SVM, ANN, and random forest classifiers. Among these three, ANN demonstrates the best performance compared to the other two and the results are shown in tables in section 5.2.2. Even though we could not find similar approaches for DFU classification in the current literature, there were similar tasks for classifying medical imagery such as skin anomaly classification, GI tract anomaly classification, etc. Moreover, we find

research that had done for binary classification of DFUs, which detect whether diabetic patients' skin imagery has an abnormality or not. Therefore compared to the literature, our approach has a novelty because we classify the DFU into 6 severity stages. The difference of the architecture of our approach compared to other medical imagery classification approaches based on transfer learning is we are using pre-trained CNNs feature extractor with GAP and SVD to optimize the accuracy and the computational time.

In order to segment and detect ulcers, we developed an automated approach for the DFU Segmentation task. Mask-RCNN model architecture was used to implement the proposed approach. The proposed approach gave promising results compared to the U-Net model in terms of performance. Overall, this solution is beneficial to both patients and clinicians since this replaces the manual wound boundary detection. A significant improvement was shown when using the Mask-RCNN model relative to the baseline model in medical image segmentation tasks as demonstrated in section 5.1.4. Therefore, this segmentation approach is applicable in other medical image datasets such as skin lesion segmentation datasets, brain tumor segmentation datasets, etc.

6.2 Future work

Other than the findings of this research, we have identified different directions worth exploring further. This study proposed an automated solution for diabetic ulcer boundary detection and segmentation. This proposed solution segment only the ulcer region and the background region, but this approach can extend to segment the ulcer lesions further. So that the clinician will be able to identify the different lesions in the ulcer and it is beneficial to do the fast treatment for patients. Moreover, the proposed automated ulcer segmentation approach can be used to develop an automatic annotating tool for medical images. The Mask Scoring R-CNN [63] or latest published segmentation models can be used in order to improve the mean average precision of the segmentation task.

The results obtained from this research work can be reused in a commercial version of healthcare application which implements for diabetic foot care. In order to achieve this, a large scale validation of this approach needs to be done. So, this will be able to approach by evaluating the proposed architectures with multiple ulcer datasets and the clinician intervention from different levels of experience.

References

- [1] S. G. Mougiakakou *et al.*, “SMARTDIAB: A communication and information technology approach for the intelligent monitoring, management and follow-up of type 1 diabetes patients,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 622–633, May 2010.
- [2] World Health Organization, *Global Report on Diabetes*. 2016.
- [3] Y. Li, L. Shen, Y. Li, and L. Shen, “Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network,” *Sensors*, vol. 18, no. 2, p. 556, Feb. 2018.
- [4] A. Ari and D. Hanbay, “Deep learning based brain tumor classification and detection system.”
- [5] A. Dutta and A. Zisserman, “The VIA Annotation Software for Images, Audio and Video,” Apr. 2019.
- [6] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2018.
- [7] T.-Y. Lin *et al.*, “Microsoft COCO: Common Objects in Context,” pp. 740–755, 2014.
- [8] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks.”
- [9] S. Kamdi and R. K. Krishna, “Image Segmentation and Region Growing Algorithm.”
- [10] A. A. Perez, A. Gonzaga, and J. M. Alves, “Segmentation and analysis of leg ulcers color images,” in *Proceedings International Workshop on Medical Imaging and Augmented Reality*, pp. 262–266.
- [11] F. Veredas, H. Mesa, and L. Morente, “Binary Tissue Classification on Wound Images With Neural Networks and Bayesian Classifiers,” *IEEE Trans. Med. Imaging*, vol. 29, no. 2, pp. 410–427, Feb. 2010.
- [12] D. Filko, R. Cupec, and E. K. Nyarko, “Detection, Reconstruction and Segmentation of Chronic Wounds Using Kinect v2 Sensor,” *Procedia Comput. Sci.*, vol. 90, pp. 151–156, Jan. 2016.
- [13] S. Saleh Al-amri and N. Kalyankar, “Image Segmentation by Using Thershod Techniques,” *J. Comput.*, vol. 2, no. 5, 2010.

- [14] S. İlkin, F. Selin Hangişi, and S. Şahin, "Comparison of Global Histogram-based Thresholding Methods that Applied on Wound Images," *Int. J. Comput. Appl.*, vol. 165, no. 9, pp. 975–8887, 2017.
- [15] L. Fraiwan, M. AlKhodari, J. Ninan, B. Mustafa, A. Saleh, and M. Ghazal, "Diabetic foot ulcer mobile detection system using smart phone thermal camera: a feasibility study," *Biomed. Eng. Online*, vol. 16, no. 1, p. 117, Dec. 2017.
- [16] G. E. Sujji, Y. V. S. Lakshmi, and G. W. Jiji, "MRI Brain Image Segmentation based on Thresholding," *Int. J. Adv. Comput. Res.*, pp. 2277–7970.
- [17] J. Umamaheswari and G. Radhamani, "An Optimal Approach for DICOM Image Segmentation Based on Fuzzy Techniques," *Adv. Comput. Sci. Eng. Appl.*, pp. 799–807, 2012.
- [18] M. K. Yadav, D. D. Manohar, G. Mukherjee, and C. Chakraborty, "Segmentation of Chronic Wound Areas by Clustering Techniques Using Selected Color Space," *J. Med. Imaging Heal. Informatics*, vol. 3, no. 1, pp. 22–29, Mar. 2013.
- [19] P. M. Azevedo-Marques, S. M. Pereira, M. A. C. Frade, and R. M. Rangayyan, "Segmentation of dermatological ulcers using clustering of color components," in *2013 26th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2013, pp. 1–4.
- [20] D. M. Dhane, V. Krishna, A. Achar, C. Bar, K. Sanyal, and C. Chakraborty, "Spectral Clustering for Unsupervised Segmentation of Lower Extremity Wound Beds Using Optical Images," *J. Med. Syst.*, vol. 40, no. 9, p. 207, Sep. 2016.
- [21] T. D. Jones and P. Plassmann, "An active contour model for measuring the area of leg ulcers," *IEEE Trans. Med. Imaging*, vol. 19, no. 12, pp. 1202–1210, 2000.
- [22] M. Kolesnik and A. Fexa, "Multi-dimensional Color Histograms for Segmentation of Wounds in Images," Springer, Berlin, Heidelberg, 2005, pp. 1014–1022.
- [23] L. Wang, P. C. Pedersen, E. Agu, D. M. Strong, and B. Tulu, "Area Determination of Diabetic Foot Ulcer Images Using a Cascaded Two-Stage SVM-Based Classification," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2098–2109, Sep. 2017.
- [24] F. Veredas, H. Mesa, and L. Morente, "Binary Tissue Classification on Wound Images With Neural Networks and Bayesian Classifiers," *IEEE Trans. Med. Imaging*, vol. 29, no. 2, pp. 410–427, Feb. 2010.

- [25] M. E. Wendelken, “Wounds Measured From Digital Photographs Using Photodigital Planimetry Software: Validation and Rater Reliability.”
- [26] “Imaging / Photography for Clinical Trials & Research | ARANZ Medical.” [Online]. Available: <https://www.aranzmedical.com/>. [Accessed: 17-Jul-2019].
- [27] “MOWA - Mobile Wound Analyzer - Wound Care Solutions (Ulcer Management).” [Online]. Available: <http://www.diabetesincontrol.com/mowa-mobile-wound-analyzer-wound-care-solutions-ulcer-management/>. [Accessed: 17-Jul-2019].
- [28] D. Filko, D. Antonic, and D. Huljev, “WITA — Application for wound analysis and management,” in *The 12th IEEE International Conference on e-Health Networking, Applications and Services*, 2010, pp. 68–73.
- [29] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.
- [30] R. Girshick, “Fast R-CNN,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.
- [31] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [32] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2018.
- [33] 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)*, 2017, pp. 618–623.
- [34] Chaghan Wang *et al.*, “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)*, 2015, pp. 2415–2418.
- [35] X. Liu, C. Wang, F. Li, X. Zhao, E. Zhu, and Y. Peng, “A framework of wound segmentation based on deep convolutional networks,” in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2017, pp. 1–7.
- [36] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for

Biomedical Image Segmentation,” May 2015.

- [37] T. Brosch, R. Tam, and for the A. D. N. Initiative, “Manifold Learning of Brain MRIs by Deep Learning,” Springer, Berlin, Heidelberg, 2013, pp. 633–640.
- [38] Y. Xie, Z. Zhang, M. Sapkota, and L. Yang, “Spatial Clockwork Recurrent Neural Network for Muscle Perimysium Segmentation.”
- [39] M. Havaei *et al.*, “Brain tumor segmentation with Deep Neural Networks,” *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.
- [40] W. Whitehead, S. Moran, B. Gaonkar, L. Macyszyn, and S. Iyer, “A deep learning approach to spine segmentation using a feed-forward chain of pixel-wise convolutional networks,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 868–871.
- [41] 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 Trans. Emerg. Top. Comput. Intell.*, pp. 1–12, 2018.
- [42] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, “The skin cancer classification using deep convolutional neural network,” *Multimed. Tools Appl.*, vol. 77, no. 8, pp. 9909–9924, Apr. 2018.
- [43] V. Thambawita *et al.*, “The Medico-Task 2018: Disease Detection in the Gastrointestinal Tract using Global Features and Deep Learning.”
- [44] M. Mateen *et al.*, “Fundus Image Classification Using VGG-19 Architecture with PCA and SVD,” *Symmetry (Basel)*, vol. 11, no. 1, p. 1, Dec. 2018.
- [45] X. Zhang *et al.*, “Gastric precancerous diseases classification using CNN with a concise model,” *PLoS One*, vol. 12, no. 9, p. e0185508, 2017.
- [46] M. Nawaz, A. A. Sewissy, and H. A. Soliman, “Multi-Class Breast Cancer Classification using Deep Learning Convolutional Neural Network,” *IJACSA) Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 6, 2018.
- [47] M. Wilmanski, C. Kreucher, and J. Lauer, “Modern approaches in deep learning for SAR ATR,” 2016, vol. 9843, p. 98430N.
- [48] H. Ravishankar *et al.*, “Understanding the Mechanisms of Deep Transfer Learning for Medical Images,” Apr. 2017.
- [49] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” Nov. 2014.

- [50] “VGG Image Annotator (VIA).” [Online]. Available: <http://www.robots.ox.ac.uk/~vgg/software/via/>. [Accessed: 08-Oct-2019].
- [51] “CVC colon DB | Machine Vision Group.” [Online]. Available: <http://mv.cvc.uab.es/projects/colon-qa/cvccolondb>. [Accessed: 08-Oct-2019].
- [52] S. Stirenko *et al.*, “Chest X-Ray Analysis of Tuberculosis by Deep Learning with Segmentation and Augmentation,” Mar. 2018.
- [53] T.-Y. Lin *et al.*, “Microsoft COCO: Common Objects in Context,” Springer, Cham, 2014, pp. 740–755.
- [54] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 2015.
- [55] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807.
- [56] C. Szegedy *et al.*, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.
- [57] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning,” Feb. 2016.
- [58] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Sep. 2014.
- [59] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 2015.
- [60] D. A. Shoieb, S. M. Youssef, and W. M. Aly, “Computer-Aided Model for Skin Diagnosis Using Deep Learning,” *J. Image Graph.*, pp. 122–129, 2016.
- [61] R. F. Mansour, “Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy,” *Biomed. Eng. Lett.*, vol. 8, no. 1, pp. 41–57, Feb. 2018.
- [62] M. Mateen, J. Wen, Nasrullah, S. Song, and Z. Huang, “Fundus Image Classification Using VGG-19 Architecture with PCA and SVD,” *Symmetry (Basel)*, vol. 11, no. 1, p. 1, Dec. 2018.
- [63] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, “Mask Scoring R-CNN,” Mar. 2019.