

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

CLASSIFICATION OF MELANOMA MALIGNANCY IN DERMATOLOGY



Ph.D. THESIS

Bilge Süheyla GAZİOĞLU

Department of Computer Engineering

Computer Engineering Programme

OCTOBER 2021

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

CLASSIFICATION OF MELANOMA MALIGNANCY IN DERMATOLOGY

Ph.D. THESIS

**Bilge Süheyla GAZİOĞLU
(504132515)**

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Prof. Dr. Mustafa E. KAMAŞAK

OCTOBER 2021

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**DERMATOLOJİDE MELANOMA MALİGNANSININ
SINIFLANDIRILMASI**

DOKTORA TEZİ

**Bilge Süheyla GAZİOĞLU
(504132515)**

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Mustafa E. KAMAŞAK

EKİM 2021

Bilge Süheyla GAZİOĞLU, a Ph.D. student of ITU Graduate School 504132515, successfully defended the dissertation entitled “CLASSIFICATION OF MELANOMA MALIGNANCY IN DERMATOLOGY”, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Prof. Dr. Mustafa E. KAMAŞAK**
Istanbul Technical University

Jury Members : **Assoc. Prof. Dr. Yusuf YASLAN**
Istanbul Technical University

.....
Prof. Dr. Muhittin GÖKMEN
MEF University

.....
Prof. Dr. Hazım Kemal EKENEL
Istanbul Technical University

.....
Prof. Dr. Aydın AKAN
Izmir University of Economics

Date of Submission : 6 September 2021

Date of Defense : 20 October 2021





*To my son,
To my spouse,
To my loving family*



FOREWORD

Above all, I would like to thank my thesis advisor Prof. Dr. Mustafa E. KAMAŞAK for his efforts, who sincerely helped and guided and supported me with his knowledge and experience while designing my research. I also thank to my examining committee members, Prof. Dr. Muhittin GÖKMEN and Assoc. Prof. Dr. Yusuf YASLAN for their considerations, valuable suggestions and crucial contributions. I am also grateful to Prof. Dr. Can BAYKAL for sharing his knowledge about dermatology.

Finally, I would like to express my thanks to all my family members for their support with endless patience and devotion at every stage of my thesis. I would like to thank my dear son, Efe, who was my motivation source with his smile and loving husband Emre for his ongoing love, support, and patience during the hard periods of my life. I am grateful to my parents and brother for support and trust me in my every aspect of my life.

October 2021

Bilge Süheyla GAZİOĞLU
(Computer Engineer, M.Sc.)

TABLE OF CONTENTS

	Page
FOREWORD.....	ix
TABLE OF CONTENTS.....	xi
ABBREVIATIONS	xiii
SYMBOLS	xv
LIST OF TABLES	xvii
LIST OF FIGURES	xix
SUMMARY	xxi
ÖZET	xxv
1. INTRODUCTION	1
1.1 Purpose and Scope of the Thesis	1
1.2 Contribution of Thesis	4
1.3 Thesis Outline.....	4
2. LITERATURE SUMMARY	7
3. DEEP LEARNING MODELS.....	19
3.1 ResNet	20
3.2 DenseNet	21
3.3 AlexNet.....	23
3.4 VGGNet.....	24
3.5 SqueezeNet	24
4. DATASET	27
4.1 ISIC- International Skin Imaging Collaboration Dataset	27
5. EFFECTS OF OBJECTS AND IMAGE QUALITY FOR SKIN LESION CLASSIFICATION	29
5.1 Methodology.....	30
5.1.1 Dataset.....	30
5.1.2 Data augmentation	33
5.1.3 Test sets with degraded images.....	33
5.1.3.1 Blurry image dataset	33
5.1.3.2 Noisy image dataset.....	34
5.1.3.3 Contrasted image dataset	34
5.2 Experimental Results.....	35
6. ENSEMBLE OF CONVOLUTIONAL NEURAL NETWORKS FOR SKIN LESION CLASSIFICATION	41
6.1 Methodology.....	42
6.1.1 Preprocessing	42
6.1.2 Training and ensemble of the models	42
6.1.2.1 Majority voting algorithm (MV)	44
6.1.2.2 Averaging probabilities algorithm (AP)	45
6.1.2.3 Weighting with confidence ratio algorithm (WCR).....	45

6.2 Experimental Results.....	45
7. DISCUSSION.....	51
7.1 Discussion about Effects of Objects and Image Quality for Skin Lesion	51
7.2 Discussion about Ensemble of Convolutional Neural Networks for Skin Lesion Classification	52
8. CONCLUSION	55
8.1 Conclusion of Effects of Objects and Image Quality for Skin Lesion	55
8.2 Conclusion of Ensemble of Convolutional Neural Networks for Skin Lesion Classification	57
REFERENCES.....	61
CURRICULUM VITAE.....	71



ABBREVIATIONS

ABCD	: Asymmetry, Border, Color, Diameter
ANN	: Artificial Neural Network
AP	: Averaging Probabilities
BCC	: Basal Cell Carcinoma
BN	: Benign Nevus
CAD	: Computer Aided Diagnosis
CAM	: Class Activation Mapping
CNN	: Convolutional Neural Network
FC	: Fully Connected
KNN	: K-Nearest Neighbour
MM	: Malignant Melanoma
MV	: Majority Voting
PCA	: Principal Component Analysis
ReLU	: Rectified Linear Unit
RF	: Random Forest
RGB	: Red Green Blue
SCC	: Squamous Cell Carcinoma
SLC	: Skin Lesion Classification
SVM	: Support Vector Machine
TDS	: Total Dermoscopic Score
TNR	: True Negative Ratio
TPR	: True Positive Ratio
WCR	: Weighting with Confidence Ratio
XAI	: Explainable Artificial Intelligence class activation mapping (CAM)



SYMBOLS

μ : Mu
 σ : Sigma





LIST OF TABLES

	Page
Table 1.1 : Melanoma statistics in Turkey [1].	2
Table 4.1 : The datasets available for dermatological studies.	27
Table 4.2 : Some studies and their accuracy values for SLC over ISIC dataset.	27
Table 5.1 : The number of images per classes in the sets.	32
Table 5.2 : The number of images per classes in the sets after data augmentation.	34
Table 5.3 : Confusion matrix.	36
Table 5.4 : Performance metrics of the models for <i>Test_Hair</i> dataset (TP: True-Predicted).	37
Table 5.5 : Performance metrics of the models for <i>Test_None</i> dataset (TP: True-Predicted).	38
Table 5.6 : Performance metrics of the models for <i>Test_Ruler</i> dataset (TP: True-Predicted).	39
Table 5.7 : Accuracy of the models for all test sets for both classes.(R: ResNet, A:AlexNet, D: DenseNet, V: VGG16).	40
Table 6.1 : Parameter values in the models.	44
Table 6.2 : Specified CNN models for every fold.	44
Table 6.3 : The model statistics of each model on each fold and ensemble of all models.	48
Table 6.4 : True predicted ratios of models.	48
Table 6.5 : Performance metrics for all models. (MV:Majority voting, AP:Averaging probabilities, WCR: Weighting with confidence ratio.)	49
Table 6.6 : A comprehensive list of skin cancer detection systems using CNN classifiers.	50
Table 8.1 : The sorted list of CNN models according to the accuracy.	58



LIST OF FIGURES

	<u>Page</u>
Figure 1.1 : Types of skin cancer [2].	1
Figure 1.2 : Sample images of melanoma(a) and basal cell carcinoma(b) and squamous cell carcinoma(c) [3].	2
Figure 1.3 : The incidence, mortality and 5-year-prevalence rates for melanoma worldwide [4].	3
Figure 3.1 : A brief CNN architecture.	20
Figure 3.2 : The difference between a residual network and a Convolutional Neural Network [5]. a: Residual network, b: Convolutional network.	21
Figure 3.3 : The architecture of DenseNet Framework [6].	22
Figure 3.4 : The difference between connectivities in networks [7].	22
Figure 3.5 : The general structure of AlexNet architecture [8].	23
Figure 3.6 : The general structure of VGGNet architecture [9].	24
Figure 3.7 : The general structure of SqueezeNet architecture.	25
Figure 4.1 : Some sample images from ISIC dataset.	28
Figure 5.1 : Block diagram of whole system.	31
Figure 5.2 : Sample images from 3 test sets.	32
Figure 5.3 : Transformed images. The images in the first row are benign lesion, in the second row are melanoma lesions.	33
Figure 5.4 : Sample images from synthetic image test sets.	35
Figure 5.5 : Graphics of metrics over noisy images.	37
Figure 5.6 : Graphics of metrics over blurry images.	38
Figure 5.7 : Graphics of metrics over contrasted images.	38
Figure 5.8 : Accuracies of deep neural network models for test sets that are calculated over 3 datasets.	39
Figure 6.1 : The overall system architecture.	43
Figure 6.2 : The statistics for each model on test set by using majority voting approach.	46
Figure 6.3 : The statistics for each model on test set by using averaging probabilities approach.	46
Figure 6.4 : The statistics for each model on test set by using weighting with confidence ratio approach.	47
Figure 6.5 : The comparison of the accuracies of three ensemble algorithms	47
Figure 6.6 : The model statistics of each model on each fold and ensemble of all models.	48
Figure 6.7 : The images classified wrongly.	50



CLASSIFICATION OF MELANOMA MALIGNANCY IN DERMATOLOGY

SUMMARY

Cancer has become one of the most common diseases all over the world in recent years. Approximately 40% of all incidences is skin cancer. The frequency of sightings of skin cancer has increased by 10 times in the last 50 years, and the risk of developing skin cancer is about 20%. Skin cancer has symptoms such as abnormal tissue growth, redness, pigmentation abnormalities and nonhealing wounds. Melanoma is a rare type of skin cancer with higher mortality compared to other types of skin cancers. Melanoma can be defined as a result of uncontrolled division and proliferation of melanocytes. Worldwide, melanoma is the 20th most common cancer and there are an estimated 287,723 new cases (1.6% of all cancers). In USA, more than two hundred thousand new cases of melanoma were diagnosed in 2021 and it increases more rapidly than other forms of cancer. Melanoma incidence increased up to 237% in the last 30 years. In our country, Turkey, melanoma is relatively rare compared to the other countries.

Cancer cells display a rapid grow and systematic spread. As in all types of cancer, early diagnosis is of great importance for the treatment of skin cancer. Early diagnosis improves treatment success and prognosis. To detect a melanoma, changes in color, shape and structure of the skin, swelling and stains on the skin are carefully examined by the physicians. Besides the physician investigation, computer aided diagnosis (CAD) mechanisms are recommended for early diagnosis.

In this thesis, deep learning models have been used to determine whether skin lesions are benign or malignant melanoma. The classification of the lesions is considered from two different points of view. In the first study, effect of objects in the image and image quality on classification performance was examined by using four different deep learning models. In addition, sensitivity of these models was tested. In the second study, it was aimed to establish a pre-diagnosis system that could help dermatologists by proposing a binary classification (benign nevi or malignant melanoma) mechanism on the ISIC dataset.

In clinical settings, it is not always possible to capture flawless skin images. Sometimes skin images can be blurry, noisy, or have low-contrast. In other cases, images can have external objects. The aim of the first study is to investigate the effects of external objects (ruler, hair) and image quality (blur, noise, contrast) using widely used Convolutional Neural Networks (CNN) models. Classification performance of frequently used ResNet50, DenseNet121, VGG16 and AlexNet models are compared.

Resilience of the mentioned models against external objects and image quality was examined.

Distortions in the images are discussed under three main headings: Blur, noise and contrast changes. For this purpose, different levels of image distortions were obtained by adjusting different parameters. Data sets were created for three different distortion types and distortion levels.

Firstly, the most common external object in skin images is hair on skin. In addition, rulers are commonly used as a scale for suspicious lesions on skin. In order to determine the effect of external objects on lesion classification, three separate test sets were created. These sets consist of images containing a ruler, hair and no external object (none). The third dataset consists only of mole (lesion) images. With the three datasets, four models were trained and their classification performances were analyzed. In fact, the best result was expected to be classified with a higher accuracy of the dataset that did not contain any object except the lesion. However, when the results are analyzed, since the image set containing hair had the highest number of images in the total dataset, the best classification performance in our system was measured by using DenseNet model on this subset. As a result of these tests, ResNet model showed a better classification performance compared to other models.

Melanoma images can be better recognized under contrast changes unlike the benign images, we recommend ResNet model whenever there is low contrast. Noise significantly degrades the performance on melanoma images and the recognition rates decrease faster compared to benign lesions in noisy set. Both classes are sensitive to blur changes. Best accuracy is obtained with DenseNet model in blurred and noisy datasets. The images contain ruler has decreased the accuracy and ResNet has better performance in this set. Hairy images have the best success rate in our system since it has the maximum number of images in total dataset. We evaluated the accuracy as 89.22% for hair set, 86% for ruler set and 88.81% for none set. We can infer that DenseNet can be used for melanoma classification with image distortions and degradations. As a general result of the first study, we can conclude that DenseNet can be used for melanoma classification since it is more resistant to image distortion.

In recent years, deep learning models with high accuracy values in computer aided diagnosis systems have been used frequently in biomedical image processing research area. Convolutional neural networks are also widely used in skin lesion classification to increase classification accuracy. In another study discussed in this thesis, five deep learning models were discussed in order to classify the images in the specially created skin lesions dataset. The dataset used in this study consists of images from ISIC dataset. In the dataset which is available in 2020, there are two classes of benign and malignant and three diagnosis consist of nevus, melanoma and unknown. We only considered images with nevus and melanoma diagnosis. Dataset had 565 melanoma and 600 benign lesion images in total. We separated the 115 images for the class of malignant melanoma and 120 images for the benign nevi class as our test set. The rest of the data was used for model training.

With pre-processing methods such as flipping and rotation, the training dataset has divided into 5 parts and the number of images in the train set was increased. DenseNet121, DenseNet161, DenseNet169, DenseNet201, ResNet18, ResNet50, VGGNet19, VGGNet16_bn, SqueezeNet1_1, SqueezeNet1_0 and AlexNet models

were trained with each subset. Using these models an ensemble system was designed. In this system, results the models were combined with the majority voting method. The accuracy of the proposed model is 95.76 % over the data set.





DERMATOLOJİDE MELANOMA MALİGNANSININ SINIFLANDIRILMASI

ÖZET

Kanser hastalığı, son yıllarda tüm dünyada en sık görülen hastalıklardan biri haline gelmiştir. Dünyada en yaygın kanser cilt kanseri olarak bilinmekte ve tüm kanserlerin yaklaşık% 40'ını oluşturmaktadır. Son 50 yılda görülme sıklığı 10 kat artan cilt kanserine yakalanma riski yaklaşık %20 olarak verilebilir. Deri kanseri, anormal doku büyümesi, kızarıklık, pigmentasyon anormallikleri ve iyileşmeyen yaralar gibi semptomlarla ortaya çıkmaktadır. Melanom nadir bir cilt kanseri türü olmasına rağmen, diğer cilt kanseri türlerine kıyasla daha yüksek ölüm oranına sahiptir. Melanom, melanositlerin kontrolsüz bölünmesi ve çoğalması sonucu tanımlanabilir. Dünya çapında, melanom en yaygın 20. kanserdir ve tahmini 287.723 yeni vaka vardır ve tüm kanserlerin yaklaşık %1.6'sı kadardır. ABD'de 2021'de iki yüz binden fazla yeni melanom vakası teşhis edilmiştir ve diğer kanser türlerinden daha hızlı artış gösterdiği belirtilmiştir. Melanom görülme sıklığı son 30 yılda %237'ye kadar yükselmiştir. Ülkemizde melanom diğer ülkelere göre nispeten daha ender görülen bir kanser türüdür Dünya Sağlık Örgütü verilerine göre, tüm kanser türleri içinde 21. sırada olduğu ifade edilmektedir.

Kanser hücreleri hızla büyür ve sistematik olarak yayılır. Kanser semptomlarının erken teşhisi, tedavi başarısını ve iyileşme oranını oldukça etkilemektedir. Tüm kanser türlerinde olduğu gibi cilt kanserinin tedavisinde de erken teşhis büyük önem taşımaktadır. Kötü huylu melanomların erken tespiti insanların hayatını kurtarabilir. Melanomu tespit etmek için cildin renk, şekil ve yapısındaki değişiklikler, ciltte oluşan şişlikler ve lekeler doktorlar tarafından dikkatle incelenmektedir ve gerekli ek tetkikler de yapılarak melanom tanısı koyulmaktadır.

Melanomun erken ve doğru teşhisi, hastalıktan kurtulma oranları üzerinde kritik öneme sahiptir. Bu amaçla son zamanlarda erken tanı için bilgisayar destekli tanı mekanizmaları önerilmektedir. Bu sistemlerde son yıllarda oldukça popüler olan derin öğrenme modelleri, potansiyel melanom lezyonlarının sınıflandırılması gibi bir çok araştırma alanında ele alınmaktadır. Bu tez çalışmasında cilt lezyonlarının iyi huylu veya kötü huylu olup olmadığının belirlenmesinde derin öğrenme modelleri kullanılmıştır. Lezyonların sınıflandırılması iki farklı bakış açısıyla ele alınmıştır. İlk çalışmada görüntüler üzerindeki dış etkenlerin sınıflandırmaya etkisi incelenmiş ve bu çalışmada dört farklı derin öğrenme modeli üzerinden sınıflandırma yapılarak bu modellerin dayanıklılığı test edilmiştir. İkinci çalışmada ise, Türk sağlık kurumundan alınan lezyon görüntüleri ile oluşturulan yeni bir veriseti üzerinde ikili sınıflandırma

mekanizması önerilerek dermatologlara yardımcı olabilecek bir ön tanı sistemi kurulması hedeflenmiştir.

Cilt görüntülerinde en sık karşılaşılan harici nesnelerden ilki, cilt üzerinde bulunan kıllardır. Bunun yanında, cilt görüntülerinde lezyonun boyutunun ölçülmesi amacıyla cetvel parçaları sonradan görüntülere dahil edilebilmektedir. Harici nesnelerin lezyon sınıflandırılması üzerine etkisinin belirlenebilmesi amacıyla da ayrı üç test seti oluşturulmuştur. Bu setler lezyon ile birlikte cetvel içeren, kıl içeren ve hiç bir şey içermeyen görüntülerden oluşmaktadır. Üçüncü veri seti sadece lezyon görüntülerinden oluşmaktadır. Aslında bu çalışmada, en iyi sonucun lezyon haricinde hiç bir nesne içermeyen veri kümesinin daha yüksek başarımla ile sınıflandırılması beklenmekteydi. Ancak sonuçlar analiz edildiğinde kıl barındıran görüntü kümesi, toplam veri setinde en yüksek görüntü sayısına sahip olduğu için sistemimizde en iyi sınıflandırma başarımla bu küme üzerinde DenseNet modeli kullanımı sonucunda ölçülmüştür. Cetvel içeren görüntülerden oluşan veri seti üzerinde yapılan testler sonucunda ise ResNet modeli diğer modellere oranla daha iyi bir sınıflandırma başarımla göstermiştir. Sistemin doğruluğu kıl içeren veri seti üzerinde %89.22 , cetvel içeren veri seti üzerinde %86 ve diğer veri seti üzerinde ise %88,81 olarak hesaplanmıştır.

Klinik ortamlarda, kusursuz cilt görüntüleri yakalamak her zaman mümkün değildir. Bazen cilt görüntüleri bulanık, gürültülü veya düşük kontrastlı olabilir ya da görüntü içerisinde ek nesneler olabilir. İlk çalışmanın amacı, yaygın olarak kullanılan evrişimsel sinir ağları modellerini kullanarak harici nesnelerin (cetvel, saç) ve görüntü kalitesinin (bulanıklık, gürültü, kontrast) etkilerini araştırmaktır. Burada, biyomedikal görüntü işleme alanında sıklıkla kullanılan 4 derin öğrenme modeli olan ResNet50, DenseNet121, VGG16 ve AlexNet ele alınmıştır. Bu modellerin sınıflandırma başarımları karşılaştırılmıştır. Modellerin hangi bozulma derecelerinde daha iyi sonuçlar vereceği incelenmiştir. Diğer bir deyişle hangi modelin hangi bozulmaya karşı daha dayanıklı veya hassas olduğunun belirlenmesi üzerine çalışılmıştır. Görüntülerdeki bozulma bulanıklık, gürültü ve kontrast değişimleri olarak üç ana başlık altında ele alınmıştır. Bu amaçla üç farklı bozulma türü için değişik parametreler kullanılarak görüntülerin bozulması sağlanmış ve her bir bozulma türü için farklı sayıda alt yapay veri setleri oluşturulmuştur. Melanom görüntüleri, iyi huylu görüntülerin aksine kontrast değişiklikleri altında daha iyi saptanabilmektedir. Ayrıca, görüntülerde kontrast farklılığı söz konusu ise cilt lezyonlarının sınıflandırılmasında ResNet modeli diğer modellerden daha iyi sonuç vermektedir. Görüntülerde sıklıkla karşılaşılan bir bozulma olayı da gürültüdür. Gürültü, melanom (kötü huylu lezyon) görüntülerinin sınıflandırılması üzerindeki performansı önemli ölçüde düşürür ve tanıma oranları gürültülü sette iyi huylu lezyonlara kıyasla azalır. Her iki sınıf da bulanıklık değişimlerine karşı duyarlıdır. Bulanık ve gürültülü örneklerin bulunduğu veri kümelerinde en iyi doğruluk DenseNet modeli ile elde edilmiştir.

Çalışmanın genel sonucu olarak, DenseNet'in görüntü bozulmalarına karşı daha dayanıklı olduğundan dolayı melanom sınıflandırması için kullanılabilmesi sonucuna varabiliriz.

Son yıllarda biyomedikal görüntü işleme araştırma alanında, bilgisayar destekli tanı sistemlerinde yüksek doğruluk değerlerine sahip derin öğrenme modellerine sıklıkla başvurulmuştur. Evrişimsel sinir ağları, sınıflandırma doğruluğunu artırmak için deri lezyonu sınıflandırmasında da yaygın olarak kullanılmaktadır. Bu tezde

ele alınan bir diğer çalışmada ise, özel olarak oluşturulmuş cilt lezyonları veri kümesindeki görüntülerin sınıflandırılabilmesi amacıyla beş derin öğrenme modeli ele alınmıştır. Bu çalışmada kullanılan veri seti, ISIC veri setinden alınan görüntülerden oluşmaktadır. 2020 yılında araştırmacıların kullanıma sunulan veri setinde benign ve malign olmak üzere iki sınıf ve nevüs, melanom ve bilinmeyen olmak üzere üç tanı bulunmaktadır. Nevüs ve melanom tanısı olan görüntüler iyi huylu ve kötü huylu sınıflarını oluşturacak şekilde ele alınmıştır. Toplamda 565 melanom ve 600 benign lezyon görüntüsü elde edilmiştir. Tüm veri seti test ve eğitim setleri olarak %20 ve %80 oranında ikiye ayrılmıştır. Test setinde malign melanom sınıfı için 115 görüntü ve benign nevi sınıfı için 120 görüntü kullanılmıştır.

Kullanılan veri seti üzerinde yüksek performans ile benign nevüs veya malign melanom cilt lezyonu tanınmaya çalışılmıştır. Eğitim veri seti 5 parçaya bölünmüş ve bazı görüntü ön işleme metodları kullanılarak alt eğitim setlerindeki görüntü sayısı artırılmıştır. Bu çalışmada her alt set için farklı bir derin öğrenme modelinin eğitilmesini amaçlanmış ve 11 farklı derin öğrenme modeli kullanılmıştır. Bu modeller DenseNet121, DenseNet161, DenseNet169, DenseNet201, ResNet18, ResNet50, VGGNet19, VGGNet16_bn, SqueezeNet1_1, SqueezeNet1_0 ve AlexNet olarak belirlenmiştir. Belirlenen her model ile herbir alt eğitim kümesi eğitilmiştir. Daha sonra, çalışmanın sonuçlanabilmesi için model sonuçlarını birleştirilmesi planlanmıştır.

CNN modellerinin birleştirilmesi, günümüzde sınıflandırma doğruluğunu artırmak için cilt lezyonu sınıflandırmasında yaygın olarak kullanılmaktadır. Bu çalışmada, bu yaklaşımı beş derin öğrenme modeli kullanarak ele aldık. Derin öğrenme modellerinin doğruluğunu ayrı ayrı artırmak istedik. Eğitilip kaydedilen her modelin test setine uygulanması sonucunda elde edilen tüm sonuçlar çoğunluk oylaması metoduyla birleştirilmiştir. Önerilen modelin doğruluğu veri seti üzerinden %95,75 olarak hesaplanmıştır. Oldukça umut verici başarımların sonuçları sistemin doktorlar tarafından bir ön tanı uygulaması olarak kullanılabileceğini göstermektedir. Bu nedenle, bu non-invaziv sistem klinisyenler için faydalı olabileceği değerlendirilmiştir.



1. INTRODUCTION

The most commonly diagnosed type of cancer is reported as skin cancer [10]. Skin cancer is formed by the uncontrolled proliferation of flat cells, basal cells and melanocytes that make up our skin. According to the American Cancer Society Foundation, one third of all cancers are skin cancer in USA [11]. Skin cancer has two main types; melanoma and non-melanoma. Figure 1.1 shows that the types and incidence of the skin cancer types in 2018 and the sample images of these types are shown in Figure1.2.

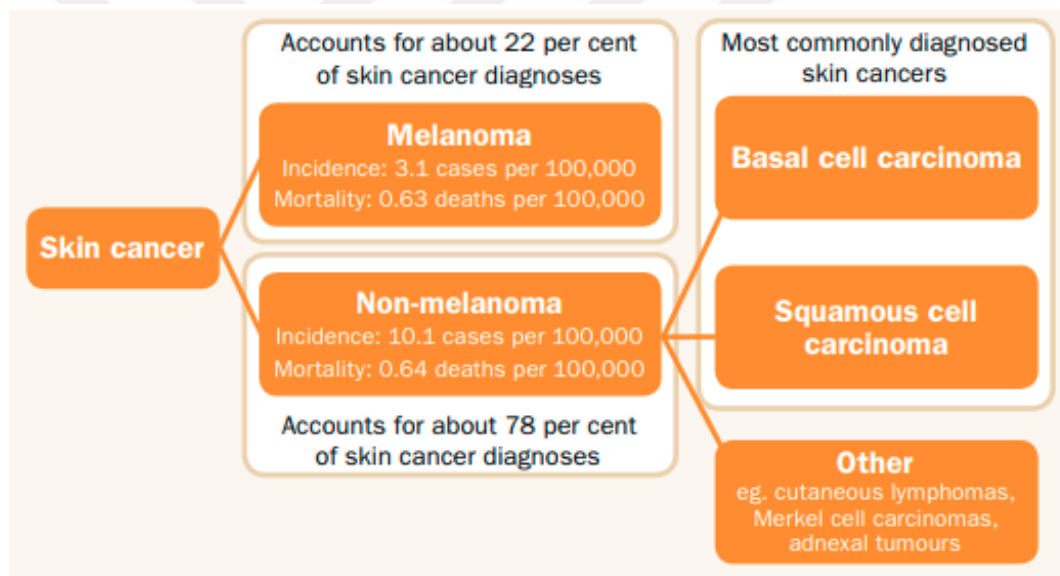


Figure 1.1 : Types of skin cancer [2].

Melanoma can be defined as a result of uncontrolled division and proliferation of melanocytes and is caused by melanocyte-stimulating cells that are pigmented on the skin.

1.1 Purpose and Scope of the Thesis

In the world, the number of new melanoma cases is reported to be 2.8 in every 100,000 individuals and 0.6 in melanoma-related deaths per 100,000 people [12]. Melanoma incidence increased more rapidly than other forms of cancer (up to 237% in the

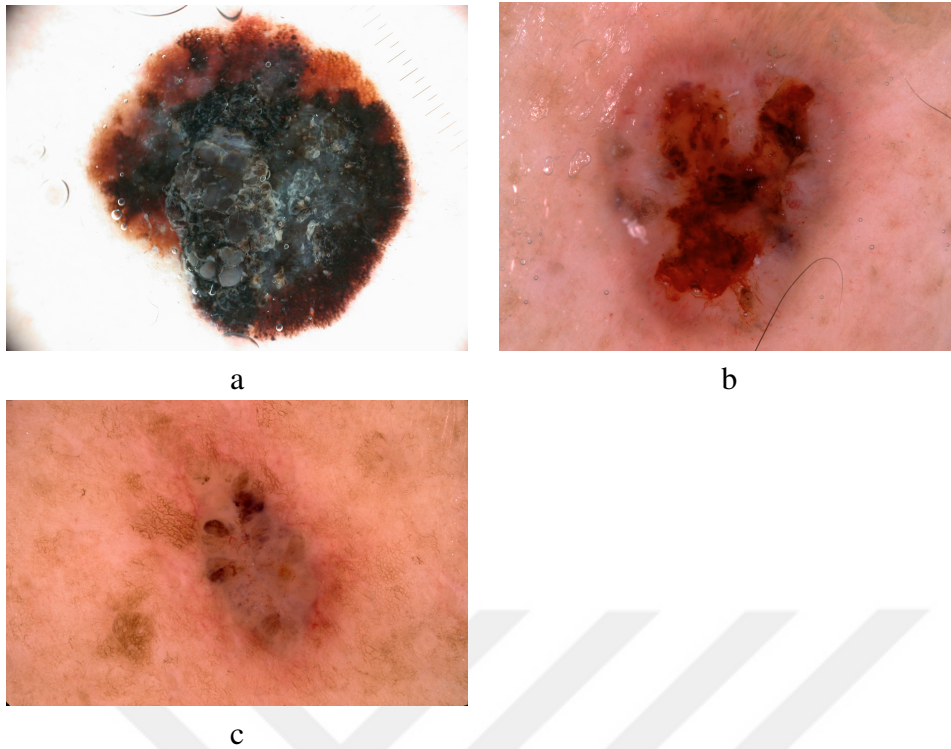


Figure 1.2 : Sample images of melanoma(a) and basal cell carcinoma(b) and squamous cell carcinoma(c) [3].

last 30 years) is reported. Melanoma, responsible for only 4% of all skin cancers, is responsible for 75% of skin cancer deaths [13]. In Europe, 62,000 new cases are detected each year [14]. In the United States (USA) in 2021, almost 100,000 melanomas were diagnosed [13]. The incidence, mortality and 5-year-prevalence rates graphics are shown in Figure 1.3.

Melanoma skin cancer incidence has also increased in Turkey in recent years, and it is in the 21st rank in all types of cancer based on the data held by World Health Organisation (WHO) [1]. 1756 new cases have been diagnosed and almost half of them has been concluded with death in 2020. These statistics are given in Table 1.3.

Table 1.1 : Melanoma statistics in Turkey [1].

	Number	Rank	(%)
New Cases	1756	21	0.75
Deaths	888	20	0.70
5-year prevalence (all ages)	5424	-	6.43 (Prop. per 100,000)

Melanoma is not so common in skin cancer types (5%), however it causes the most death. Despite high mortality rates of melanoma, early detection of malignant melanomas can save human life. To detect a melanoma, changes in color, shape and

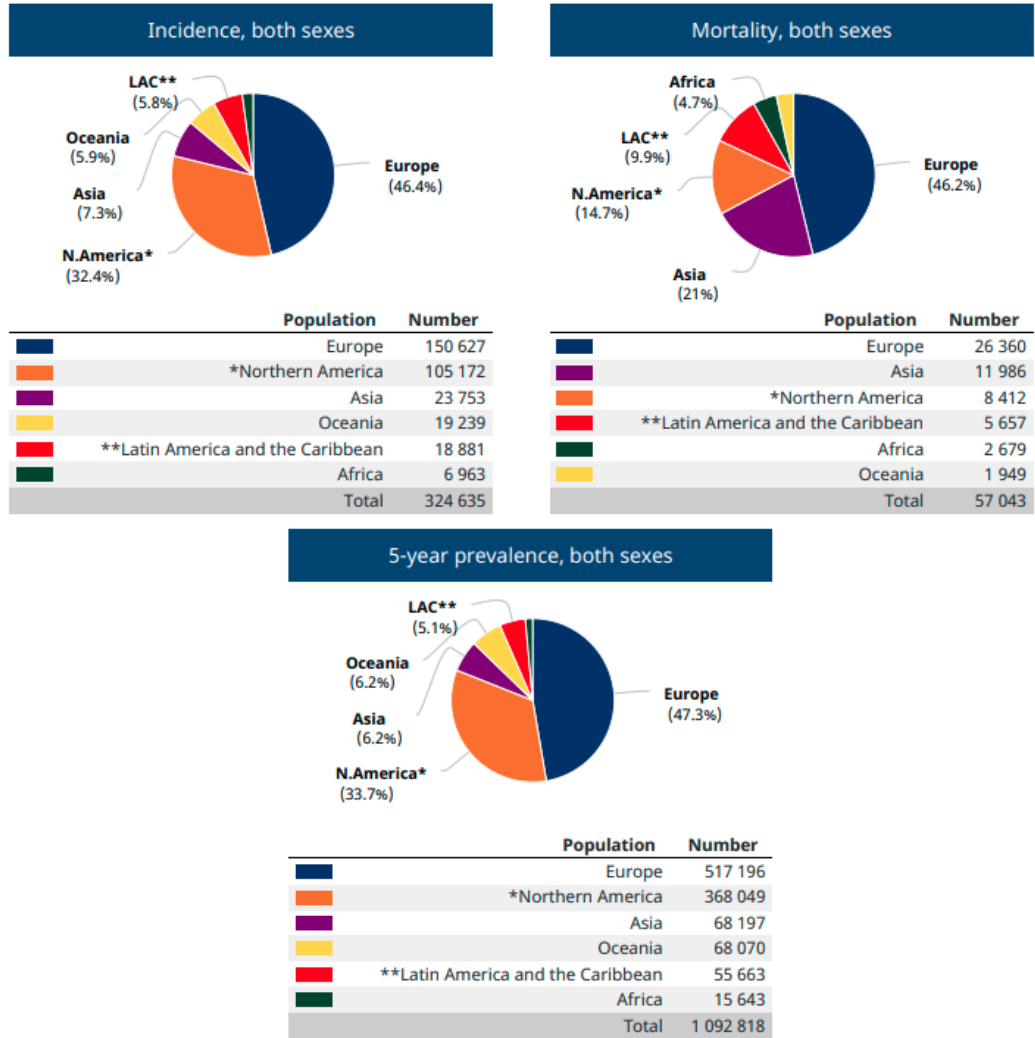


Figure 1.3 : The incidence, mortality and 5-year-prevalence rates for melanoma worldwide [4].

structure of the skin, swelling and stains on the skin are carefully examined by the doctors. Melanoma can be cured by treatment in early stages, but tends to spread if not caught early. Melanoma is rare but more serious than other skin cancers.

For early detection of malignant melanomas, a computer aided diagnosis (CAD) systems can be useful. CAD systems for early melanoma detection are used in many academic studies. In these studies raw RGB images of moles are taken and preprocessed. After that, features are extracted and machine learning algorithms are applied. A CAD system with these stages has effectively good results.

The purpose and scope of the thesis is to speed up the diagnosis process of melanoma malignancy. The proposed CAD system can be integrated to the systems used in hospitals and used as a pre-diagnosis tool for clinicians.

1.2 Contribution of Thesis

We can summarize the contribution of thesis on skin lesion classification problem with two different ways.

We first analyzed how the classification can be affected by additional objects or image degradations. Presence of hairs on the skin can often and naturally be described as extra objects in lesion images. In addition, there can be ruler in the images since the dermatologists use them as a reference in order to measure the size of the lesions. Other factors such as contrast changes, noise and blur in the images can be expressed as image distortions that cause difficulties in image processing. We have investigated the effects of these image distortions for classification problem. For this aim, we examined all the pictures one by one and separated the images with ruler or hairs in ISIC dataset. In addition, we created degraded image dataset for different levels of blur, noise and contrast changes. First part of the thesis analyzed the effects of additional objects and image distortions in skin lesion classification with different deep learning models in the literature [15].

The other contribution of this thesis is to create a CAD system for classification of skin lesions. This system is planned to be helped in early diagnosis of skin cancer. The ISIC dataset for challenge 2020 is used for this system. In this study, the dataset was divided into parts with a K fold cross validation style approach and train each fold with a different CNN model. The final result was obtained by fusing the results all CNN models. This resulted in a CAD system with higher accuracy.

1.3 Thesis Outline

The rest of this thesis is structured as follows: In Chapter 2, we reviewed the existing studies on skin lesion classification. In Chapter 3, we summarized the deep learning models used in this thesis. In Chapter 4, we provide details of the publicly available dataset, ISIC, which is commonly used dataset on skin lesion classification problems including this thesis. The effects of additional objects and image degradations on melanoma classification and results of four different deep learning CNN models are provided Chapter 5. Details of the proposed CAD system and its performance are

described in Chapter 6. Finally, findings and results of the thesis are given in Chapter 7 and 8 respectively.





2. LITERATURE SUMMARY

Studies of skin lesion classification consist of many different components such as preprocessing, segmentation and classification. In melanoma classification, an important preprocessing step is hair removal [16–18]. Hair on the dermoscopic images may block the extraction of proper features from the image because it closes the main structures in the image. Hence, hair need to be removed before creating a loss in the images. After hair removal, many studies dealt with the segmentation of skin lesion for better image understanding [19,20]. Segmentation aims to differentiate the foreground and background in the skin images. Foreground includes the lesion and background includes the skin and various textures that make the classification harder. With segmentation, the skin lesion can be processed directly without any artificial affects. Segmentation process can continue with classification or feature extraction step. Features are extracted from skin lesions and they can be in many different forms such as ABCD features (asymmetry, border, color and diameter) and other image features [19, 21]. Final step is classification. Deep learning [22, 23] and machine learning algorithms [24,25] are very popular for classification problems. Deep learning algorithms consist of convolutional neural networks (CNN) and machine learning algorithms such as support vector machines (SVM), K-nearest neighbour (KNN) and random forest (RF) are mostly used in skin lesion classification (SLC). Some articles about skin lesion classification are summarized below.

Toossi et.al. proposed a two layered system for hair removal in dermoscopic images [17]. In the first step, they dealt with the hair detection. Firstly, principal component analysis (PCA) was applied to the images to enhance the image contrast and to convert red green blue (RGB) image to grayscale image. After that, for noise reduction a 3x3 sized Wiener filter was applied. Canny edge detector was used for the detection of the edges of the hairs and segmentation of the hairs was succeeded with dilation morphological operator. Dilation and filling operators were used for smoothing hair lines in the images. As the second step in the system was hair removal was acheived

by inpainting. In the study, multi-resolution coherence transport inpainting method was proposed. It was the combination of wavelet and simple coherence inpainting methods. Experiments were tested on 50 images and true detection rate was measured as 93.2% and false positive rate was 4% as overall system performance.

A system proposed by Korjakowska and Tadeusiewicz had 4 stages [16]: First step was black frame removal. The lightness value of the pixels were calculated and if this value was less than 15, then it was determined as black pixel. In addition, if more than half of the pixels in a row were black, that line was labeled as a black frame and eliminated. As a second step, they applied Gaussian filter for noise reduction. Next stage was hair removal. They analyzed that top-hat transformations could be used for this aim. In this approach, RGB images were converted to grayscale images. Negative Laplacian filter was applied to increase the contrast. Hair were detected with black top hat transform which is a different form of the closing operator. The last step of the top-hat transformation was inpainting to find the whole hair and reconstruct the blanks of the removed hairs. Diagnostic accuracy and sensitivity were the metrics to measure the system performance and they were measured as 88.7% and 90.8% respectively.

In another study, a combined algorithm consisting of watershed marker controlled and canny edge detector methods was proposed [19]. There were three steps in this study: image segmentation, feature extraction and classification. In first step, two methods were used for segmentation: magnitude gradient algorithm and canny edge detector. In next step, asymmetry, border, color and diameter features are extracted according to the ABCD rule. They calculated a total dermoscopic score (TDS) value from these four features. According to these TDS values of images, classification was performed. They tested the proposed system with 10 images and performance of the system was 90% for melanoma detection and 80% for non-melanoma detection.

In another study, a smart phone application was developed for detection of melanomas [20]. System was designed in a client-server architecture. A user takes a photo of the mole image via a smart-phone and this image was sent to the server. In server side, feature extraction and classification were performed and compared with the images in the dataset. The classification result was returned to the user's smart-phone. The framework included 3 different steps which were preprocessing, segmentation and classification. The first step had three parts inside. With black frame removal,

smoothing and black hair inpainting algorithms, air bubbles, light and dark hairs and black frames according to the HSL space were removed. The seeded region growing algorithm was applied for segmentation step. After that, two classification methods- micro-melanoma algorithm and specific melanocytic lesion algorithm- were used in the last step. The diameter of segmented image of the skin moles was measured by the user while sending the image and if it was less than 5 mm, then it was classified with micro-melanoma algorithm. Otherwise specific melanocytic lesion algorithm was used. Sensitivity, specificity and AUC metrics were used to assess the system performance. Tests were performed on 144 images and the overall results were 92.2%, 97.2%, and 96.6% as TPR, TNR and AUC respectively.

In a study by Salido and Ruiz, deep learning model was trained for melanoma detection [18]. Their system had three parts: Hair removal, lesion segmentation and classification with CNN. In hair removal step, they used median filtering and morphological operators as hair detection. Bottom-hat filtering was applied to the images obtained by median filtering for the hair detection aim, then they applied a closing operator to remove small objects in the images. After detecting hairs, they used harmonic inpainting algorithm to fill the removed hair pixels with the new values. PH² dataset was used for this problem in the study. After image processing steps, transfer learning and classification were performed. For transfer learning, they used AlexNet pretrained network. There were 96 kernels, 48 of them were color agnostic and others are color specific. There were two classes - melanoma and non-melanoma. The classification was done with AlexNet. Hyper parameters for the used network were; learning rate is 0.001; maximum epochs were 1000 and mini batch size was 64. They used stochastic gradient descent algorithm and momentum parameters were used to train the CNN. The overall system accuracy was 92.5% as testing result and 95% as validation result.

Two different deep learning networks was used in a study for melanoma detection [26]. They used ISIC 2017 dataset. Preprocess was applied for cropping and resizing the images. After that, data augmentation was performed by rotating and flipping over the rotated images and they obtained a new dataset. Fully convolutional residual networks(FCRN) was applied to each dataset namely original and flipped datasets separately. FCRNs were used to for both segmentation and classification

simultaneously. The proposed framework was named as lesion indexing network. For feature extraction, they proposed a lesion feature network (LFN) framework. Superpixel extraction was used in this part. Each superpixel area was categorized as pigment network, negative network, streaks, milia-like cysts and background. The numbers of these patches were in different ranges. Hence, some operations like rotation were used to equalize the numbers. LFN could be considered as a CNN which included 12 convolutional layers. Stochastic gradient descent algorithm was used for optimization of the network. After validation and testing, they obtained the accuracies for each tasks: 75.3% for segmentation, 84.8% for feature extraction and 91.2% for classification.

Albahar proposed a novel regularizer for classification of skin lesions [27]. The regularizer was embedded in a deep convolutional neural network. In preprocessing steps, images were resized and gamma correction was applied. After preprocessing, the images were fed to CNN. It had 2 convolutional layers with 32 and 64 filters respectively. The regularizer was embedded to each layer of CNN. Max pooling and dropout layers were applied and final layer which was fully connected layer with 128 neurons was applied the images. 5600 training and 2400 validation images were used in this architecture. The system was trained with 100 epochs and 97.49% accuracy was obtained.

Melanoma recognition study focused on the aggregated deep convolutional features of the dermoscopy images [22]. As in other studies, there was a preprocessing step in this study. Firstly, images were resized and normalized. Normalization procedure was done with subtraction of mean values of all images from the original image. After that, rotation and translation were applied the images. The next step was extraction of local convolutional features of the images from the pretrained network. After preprocessing operations, 4 augmented images (per image) were obtained which are then passed to CNN. Spatial map with width, height and depth parameters were possessed for each augmented image and four of them were combined with Fisher vector encoding (GMM and PCA were used) and deep features of the image was obtained. SVM classifier was used for the classification. They used stochastic dual coordinate ascent algorithm, it is used to minimize the regularized loss with the fast convergence property during SVM. ISIC dataset was also used in this study. The number of images training

and test set were 900 and 379 respectively. In evaluation, they tried four different networks: ResNet50, ResNet101, ALexNet, and VGGNet. They got the best results with ResNet50 and Deep CNN-FV fusion algorithm. Their overall accuracy was 86.81%.

Attention residual learning for skin lesion classification was a study that fuse the attention residual learning and convolutional neural networks [23]. In this study, there were three sections: Attention residual learning (ARL) blocks, global average pooling and classification. ARL was a combination of residual and attention learning. Residual function took the input data of CNN and this input data was added to the function and calculated a new function at the end of the CNN. They used attention learning to have more powerful representation of network by producing the attention weights of the already trained CNN. For this aim, there was no need to add extra layers to learn. They emphasized the higher layer, so they used the feature maps of higher layers as the mask of the lower layers. To combine these two learning into an ARL block. They added the residual function, attention map and the input as output of the CNN. They implemented systems with two different network layers (14 and 50). ISIC dataset was used for evaluation; 2000 images for training, 150 images for validation, 600 images for test. After testing many combinations, they obtained the best accuracy with ARL-CNN50, it was calculated as 85% for melanoma classification and 86.8% for seborrheic keratosis classification.

Nie et.al collected a dataset from University of Salerno and they had 760 training images and 191 test images [28]. They divided the training set into five folds and trained 5 different VGG16Net models. Four of these models were used for training and 1 fold was reserved for validation. After training processes, they tested their system over 5 models and the result is averaged over them. The accuracy of their system for test set was measured as 63.35 % and 68.82 % for validation sets.

In the study about melanoma classification, he handled with 4 pretrained and 1 proposed CNN models also some machine learning algorithm such as SVM and K-NN [21]. 3920 images were chosen from ISIC dataset. Firstly, feature extraction and Haralick texture features were used were extracted. After that, pretrained models (AlexNet,VGGNet, ResNet, and GoogLeNet), and a proposed model (C4Net) were trained. These 5 models were compared and best accuracy was obtained with C4Net.

It contained 4 convolution, 3 max pooling, 2 fully connected layers and one output layer. Dataset was divided into 10 folds and it was calculated over 150 epochs. The average accuracy was calculated as 96.94 %.

Acosta's study was melanoma diagnosis using deep learning techniques on dermoscopic images [29]. They proposed two-stage-algorithm. Firstly, they used Mask R_CNN model for segmentation and in the next stage they handled with classification of skin lesions with ResNet152. Mask R_CNN model specified the pixels that can be lesion pixels which they named them as candidate pixels. The probabilities of these candidate pixels being a lesion were calculated. The region was bounded and cropped according to the probability values. For this aim, they used images and masks in the ISIC 2017 dataset. Then the cropped images were used in classification. Transfer learning was applied for classification and model was trained for fine tuning. They reached the classification accuracy of 84.8 %.

Maiti and Chatterjee aimed to improve detection of melanoma and nevus with deep neural networks [30]. The proposed system consists of 4 modules: image preprocessing, feature extraction, data preprocessing and deep learning. Image denoising and contrast enhancement are applied as preprocessing. For denoising, they use an-isotropic diffusion function. It preserved edges and made blurry background in the image. By modifying an-isotropic diffusion function, they normalised images. Next step was contrast enhancement which was achieved with modified an-isotropic diffusion with sigmoid functional. After that, processed images were sent to the lesion mask generation module in which segmentation part was performed. Feature extraction module included 2 main feature selection methods. They got 22 features with Haralick co-occurrence matrices and 12 features with shape feature algorithm. After that, polynomial features with size of 629 was produced from 34 features and new set was reduced by principal component analysis (PCA) and sent to deep neural network. MedNode dataset with 170 images and ISIC dataset with 2000 images were used in their system. They used 12 machine learning algorithms and deep neural networks for classification. Their best results were obtained with deep learning. Overall system performance was reported as 96.8%.

Another study included three different deep learning method fusion [31]. As preprocessing, they first normalised the images by subtracting mean values. Also,

images were resized with bicubic interpolation method. Final preprocessing was augmentation including rotation and horizontal flipping. ISIC dataset was also used in this study and the number of dataset was increased by 8 times. After these operations, images were sent to three neural networks: VGG16, AlexNet and ResNet18. Networks were used for feature extraction from the images. Features were extracted from the last fully connected layers of the neural networks. Extracted features were the inputs of non-linear multi-class support vector machine classifier. There were 3 classes in the dataset and SVM was set for malignant melanoma vs. all and seborrheic keratosis vs. all. Three SVM classifiers were implemented for each network and their results were fused and evaluated by probabilities using logistic regression function. After many experiments, they possessed the best AUC value with the fusion of all three networks. They calculated the overall system performance as 90.69 %.

A team participating ISBI challenge 2018 published a study about transfer learning on dermoscopic images [32]. They used the ISIC dataset with 7 classes but amount of data in each class was different from each other. Firstly, the normalization preprocessing step was applied. Max RGB normalization algorithm was applied the images. After that, by horizontally flipping the data was doubled and by rotation data amount in the classes was balanced. They obtained almost 13000 data for each class. The next step was training the network. VGG16 and GoogLeNet were chosen to use in transfer learning. Last three convolutional layer was trained and weights are learned. As a result, they validated their system with ISIC validation set including 193 images and the validation accuracy was calculated as 79.7 % for VGG16 and 80.1 % for GoogLeNet.

In the study of team of Minagawa and Koga, the first step was normalization of the images [33]. Intensity and color balance values of the images were normalized. After that, these images were sent to two classifiers. One classifier was used for training Seborrheic Keratosis (SK) versus others and the other one was used for Malignant Melanoma (MM) versus rest. They performed rotation, scaling and translation on the images. They used a 50 layer ResNet in Keras as CNN with modifications. It was analyzed that age and sex information were very important features to separate SK from the others. Hence, they added these features for SK classification. There was no lesion segmentation or cropping algorithm implemented. They used 783 MM images,

320 SK images and 2341 Nevo Cellular Nevus (NCN) images for training of CNNs. As a further study, the RMSProp and AdaGrad optimization methods were used for finding the best combinations of fine-tuned CNNs. The AUC measure was used to show the results and their results are 0.924 for MM classification and 0.993 for SK classification.

Ivan Gonzalez-Diaz proposed a system to classify the nevus, melanoma and seborrheic keratosis [34]. CNN was also used in this study. They used a pipeline for classification. First process of the pipeline was lesion segmentation network. It segmented the image and outline the area of the lesion and generated a mask with the help of the fully convolutional networks. The mask and the original image were sent to the data augmentation module. In this module, the images were rotated and new set of 24 rotated images was obtained. In the next stage, structural segmentation network, they extracted 8 features which were important structures for dermatologists to diagnose. The extracted structures were labeled by physicians. According to the labels, it was associated with the probabilities. Final module of the pipeline was diagnosis network. ResNet-50 was used as basis. They implemented a 3-layer-system for final layer in the classification.

Valle et. al. developed a deep learning solution to the problem [35]. They found 3 bottleneck in the system. One of them was the limited number of images. They solved this problem by using 6 different datasets. ResNet-101 and Inception-V4 models were used and the models' experiments were done with Slim/Tensorflow. More images to train and more layers (deeper) in the models had more successful results. Augmentation and normalization of the images were the important methods to increase the success rate they analyzed. They added a SVM classifier as an extra layer at the end of the network. Their overall system AUC was 0.908.

Xie et al. used an algorithm to generate high resolution feature maps which included spatial and discriminative feature of boundaries and obtained segmented lesions with the success rate of 85.7% on PH² dataset by using FCN-8 and U-Net CNN models [36]. In another segmentation study by Goyal et al. an ensemble of two different CNN models -Mask RCNN and Deep LabV3- was developed and tested on ISIC datasets. They obtained 93.8% segmentation accuracy [37]. CNN based models were used very commonly for skin lesion segmentation [38,39]. In a study by Jafari et al., skin images

were divided into patches that were sent to a CNN model. CNN model extracted some local and global features from patches and detects the borders of lesion. Their best accuracy was 98.7% [38]. Yuan worked with a few samples of skin images for segmentation and FCN were used in their study. They took the problem as pixel-wise classification instead of segmentation and proposed a novel loss function that uses Jacard distance, they achieved the best accuracy of 95.5 % [39].

Besides segmentation deep learning algorithms become very popular in the classification of lesions [23, 27, 40]. In addition to these studies, many systematic reviews about lesion classification problem were recently published [41–43]. In these three review studies, performances of both traditional classifiers (ie. SVM, KNN, Naive Bayes and Decision Trees) and CNN model were investigated for classification of melanoma.

Training of deep learning models require large amounts of labeled data. When the amount of labeled data is low, models that are trained on other types of images are used as a baseline model. These models are then fine tuned for the specific purpose using transfer learning methods. In Harangi ensembles of DNN were used for skin lesion classification [44]. In this study, ISBI 2017 Challenge dataset was used with three classes: nevus, melanoma and seborrheic keratosis. Training images were augmented by rotation, flipping and cropping operations. They recommended to use of fusing four different CNN models: AlexNet, ResNet, VGGNet and GoogLeNet. The fusion of the outputs on the classification layer was handled and the average accuracy was calculated as 86.6% after fusion model with weighting.

In another study by Hosny et al., classification of skin lesion were performed using transfer learning and AlexNet [45]. They used three different datasets, which are MedNode, DermIS- DermQuest and 2017 ISIC Challenge. Their augmentation technique was based on image rotation. They rotated images in two ways. One way was arbitrary rotation and the other was systematic rotation (between 0° - 360° with regular intervals of 5°). They used modified AlexNet for transfer learning, and changed the classification layer with softmax layer. Stochastic gradient Descent algorithm was used with a small learning rate to update weights of the CNN model. Their system performance was measured as 96.9%, 97.7%, 95.9% for Derm, MedNode and ISIC datasets respectively.

Mahbod et al. fused deep features for classification of skin lesion [46]. In this study, ISIC Challenge 2017 dataset was used. First steps in their pipeline were image preprocessing, image normalization, and resizing. After that, four CNN models; AlexNet, VGG16, ResNet18 and ResNet101, were trained for feature extraction. The features were used as the outputs of fully connected (FC) layers of the fine tuning networks. They added two FC layers to the networks as replacement of the last FC and output layer. After extracting features, SVM classifier were trained for three classes. The best performance was obtained with fusion of all fine-tuned networks. The accuracy was 87.26% and 95.52% for melanoma and seborrheic keratosis respectively.

Perez et al. used a CNN model for melanoma classification [47]. In this study, many experiments were conducted to demonstrate the performance of different CNN architectures for melanoma classification. They also used ISIC Challenge 2017 dataset and 9 different CNN models. The models were DenseNet, Dual Path Nets, Inception-V4, Inception-ResNet-v2, MobileNetV2, PNASNet, ResNet, SENet, and Xception. They performed the experiments 3 times on 5 test splits with 9 architectures. They experimented with architectures of the models such as initialization of last layer, augmentation, dropouts, data shuffling. These experiments revealed how they effect the system performance.

A study based on the new trends such as transfer learning, deep learning and IoT system for skin lesion classification was proposed and uses the ISIC and PH^2 datasets [48]. The images from two datasets were sent to the CNN models for feature extraction. They extracted features using 12 different CNN models. After that, they used five different machine learning algorithms for classification of extracted features. They compared the results of CNN model and other classification methods. The best accuracies were obtained with combination of DenseNet201 and KNN algorithm, calculated as 96.805% for ISIC dataset and 93.167% for PH^2 dataset.

In another study, melanoma and benign nevus lesions were tried to be classified [49]. The researchers applied three different classification methods; MLP, CNN(ConvNet), CapsNet to the skin lesions. They used MLP with 3 layers and 2352 input neurons. CNN model had 2 convolutional, pooling and fully connected layers, it was 8-layer-CNN model. In CapsuleNet, after ReLu layer, there is a primary capsule layer. The capsule is a kind of vector which has size and direction information and it keeps

the object feature. They used a dynamic routing by agreement method instead of max pooling layer. Their dataset consisted of 170 images including 70 melanoma and 100 nevus samples. The accuracy was 76.4% with MLP, 95.1% with CNN and 98.9% with CapsNet.

Barata et. al. proposed a CAD system including segmentation, feature extraction and classification processes for diagnosis in skin lesion [50]. In that system, they dealt with some problems. One problem was about the dataset that did not contain sufficient information. Also, difficulty in identifying features and difficulty in diagnosis of selected features were considered other issues. They used machine learning algorithms for this aim. They used SVM, random forest and KNN machine learning algorithms for classification and their results were fused with logistic regression and median rule approaches. The EDRA atlas dataset with 804 images was used. The best accuracy was obtained with SVM with RBF kernel and calculated as 85.9%.

Shorfuzzaman stacked the five different CNN models for skin lesion classification problem: ResNet50, VGG-16, DenseNet121, Xception, and EfficientNetB0 [51]. In this study, ISIC archive images used and they were applied some preprocessing operations such as noise reduction, resizing, normalization and enhancement. After that, they trained the processed images with five CNN models and took three best performing models. DenseNet, Xception, and EfficientNet were selected for model evaluation and three models produced three probability vectors. According to the combination of these vectors the FC layer with sigmoid function was used as meta learner and in this way, classification was handled. The proposed method was achieved the accuracy with the rate of 95.76%.



3. DEEP LEARNING MODELS

Studies on deep learning have gained momentum in recent years with increasing hardware availability. DCNN (Deep Convolutional Neural Network) becomes very popular in image classification, scene classification, character recognition, image segmentation etc. In almost all of these areas, the most successful results were obtained with DCNN. Due to its performance, we also preferred to use DCNN within the scope of this thesis.

A DCNN architecture is an architecture that consists of one or more convolution layers (usually one sub-sampling layer) and one (or more) fully connected multi-layer artificial neural networks. DCNN architecture is designed to take advantage of the 2D structure of input images. The neurons in the hidden layer are connected only by neurons in a certain $k \times k$ zone, ensuring that the network is simple. Furthermore, since shared weights are used, the number of parameters is much less than a fully connected artificial neural network.

Deep learning has been the focus of attention in recent years and it is one of the most commonly used methods in the biomedical image analysis. It allows us to train a model to predict outputs with a given data set. For this aim, deep architectures which are based on neural networks (residual, convolutional neural networks etc.) are used. There are input, hidden and output layers in neural networks and many hidden layers. Convolutional neural networks (CNN) are form of deep learning architectures. In cNN architectures, the input is fed to the convolution layers whose output is connected to the pooling layer. Deeper layers can be obtained by using these two layers repeatedly. After convolution and pooling layers, fully connected (FC) layers are used. FC layers also can be repeated. Finally, a fully connected layer with an activation function such as rectified linear unit (ReLU) or SoftMax is applied and output is produced. The concept of CNN models are shown in Figure 3.1.

In the literature, many networks have been designed. Some of these are presented to use publicly such as ResNet, DenseNet, etc. These models are implemented and

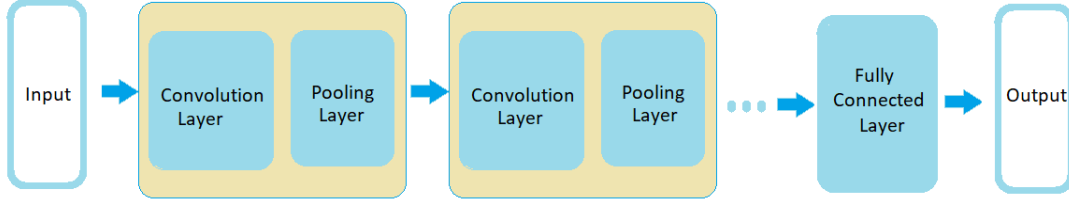


Figure 3.1 : A brief CNN architecture.

trained for a specific problem and all weights are saved. They are called as pretrained models. The pretrained models can be directly used for other problems. We also used pretrained models in this thesis as it requires less data to train them. They are explained in next section.

3.1 ResNet

ResNet is one of the most commonly used network in deep learning studies. It is a network proposed by He Kaiming et al. in 2015, ranking first in the ImageNet Large Scale Visual Recognition Challenge-2015 classification challenge [52]. ResNet is an abbreviation of Residual Network. It is based on the residual learning. ResNet also managed to reduce the number of parameters, which is an important factor in computational complexity.

ResNet model has a different structure from the traditional sequential network architectures. ResNet differs from other architectures with its micro architectural module structure. Changes between some layers ignored and it can be preferred to switch to the lower layer. With this property, ResNet architecture has improved performance. In addition to this layer structure, we can find the information about the connections between layers. The difference between a residual network and a convolutional neural network is shown in Figure 3.2.

Residual networks use a residual mapping and The part of the network from the input to the output can be mapped with a nonlinear $H(x)$ function. In ResNet architecture, this path is mapped to another nonlinear function defined as $F(x) = H(x) - x$ instead of $H(x)$. In addition, by making a shortcut connection from input to output, the x (input) value is added to the $F(x)$ function arithmetically. Then, the function $F(x) + x$ and x are passed to ReLU together. Therefore, by adding input to the end of second layer, it is aimed to transmit the values from the previous layers to the next layers more strongly.

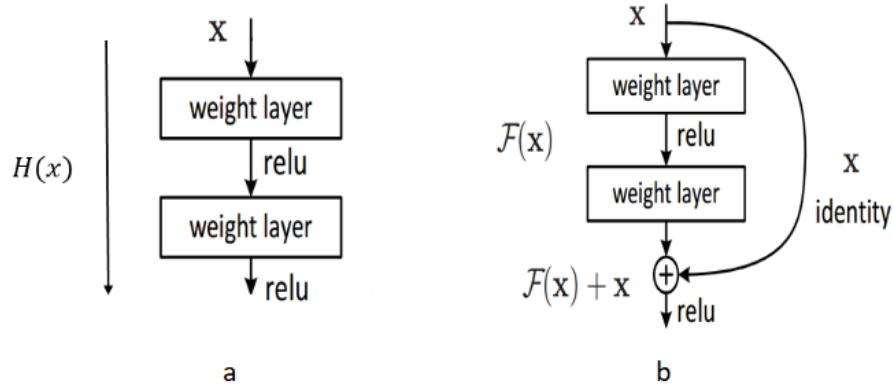


Figure 3.2 : The difference between a residual network and a Convolutional Neural Network [5]. a: Residual network, b: Convolutional network.

As stated in the article published in 2015, 152-layer ResNet was the deepest network feature ever run on ImageNet. ResNet also has fewer parameters than some other deep networks. This plays an important role in training faster and with less data. There are many variations of ResNet: ResNet50, ResNet101, ResNet152, etc. In this thesis, we used different ResNet models with different number of layers.

3.2 DenseNet

DenseNet can be summarized as dense transform network that connects each layer to all other layers in the feed forward style. It was developed by Huang et al. [53].

DenseNet uses less parameters compared to other architectures as it aims to reuse the feature map produced by previous layers. Hence, a k -layer does not need to learn that map over again. That layer has access to features from other layers. Due to this feature, DenseNet is compact and effective network. DenseNet layers are very narrow (for example, 12 feature maps per layer), it adds a small set of feature maps to the network's collective information and keeps the remaining feature maps unchanged. The final classifier decides the class of input based on all feature maps on the network. One advantage of DenseNet is advanced information and gradient flows that makes it easier to train across the network. Each layer has direct access to gradients from the loss function and the original input signal, which leads to an implicit deep control. This helps the training of deeper network architectures. Furthermore, intensive connections have a regular effect that reduces excessive compliance with tasks with smaller training set sizes. The framework of DenseNet model is shown in Figure 3.3. In the figure there

are 3 dense blocks with 16 convolutional layers. DenseNet is a type of convolutional

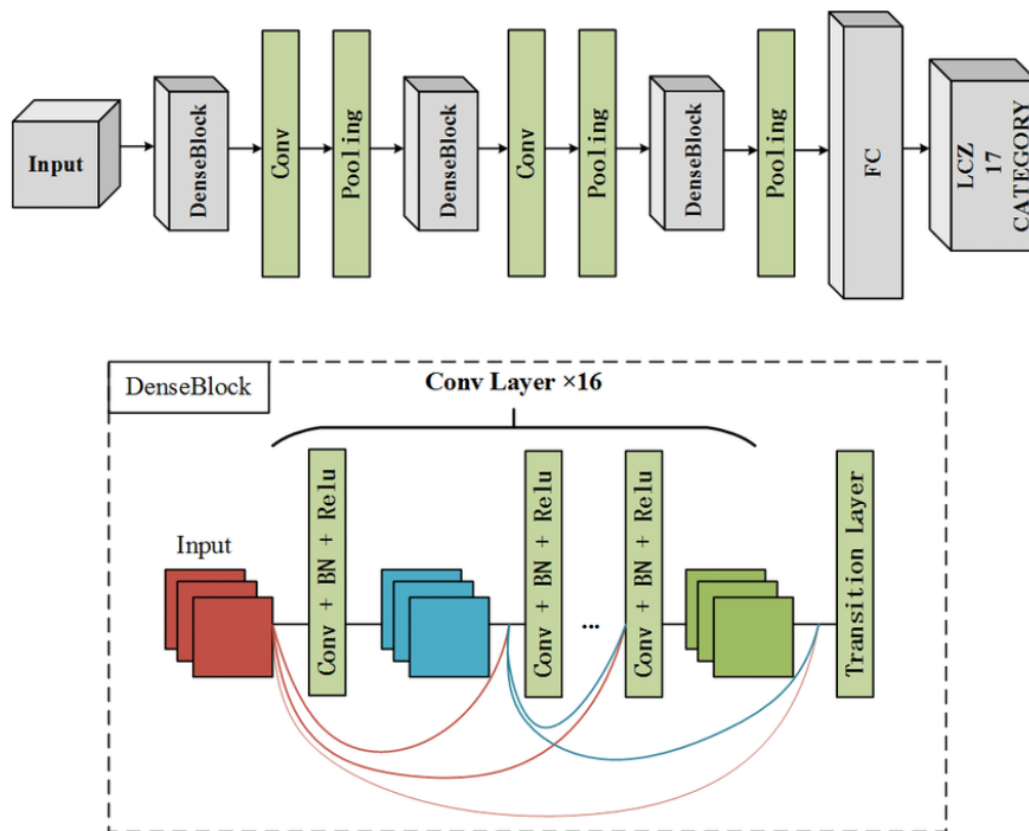


Figure 3.3 : The architecture of DenseNet Framework [6].

neural networks. In DenseNet, there are links that connects all layer to each other. Hence the features can be reused in other layers, which is the difference of DenseNet and ResNet. The difference between networks can be seen in Figure3.4.

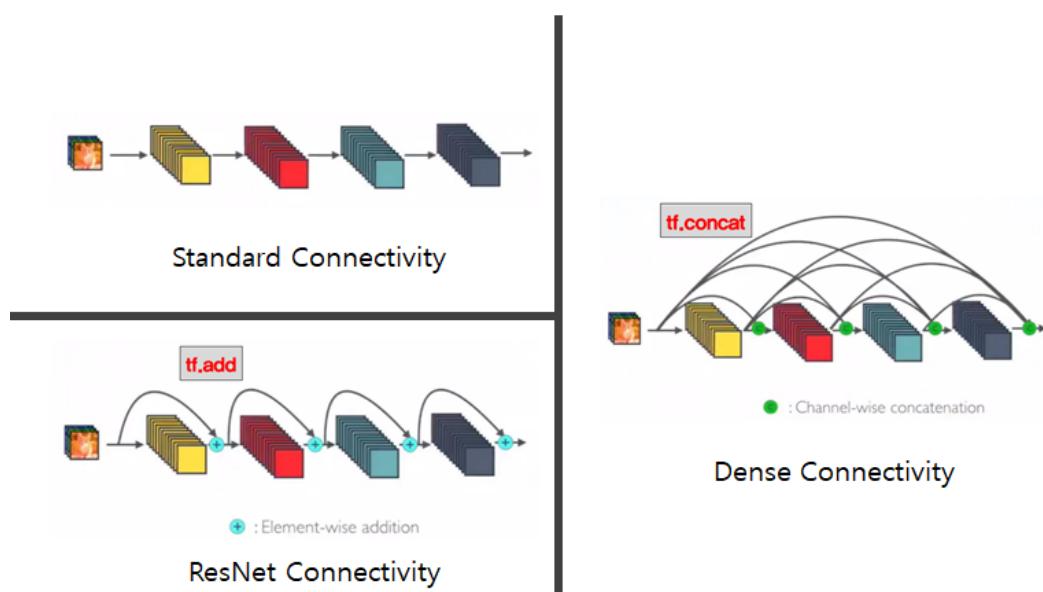


Figure 3.4 : The difference between connectivities in networks [7].

3.3 AlexNet

It is one of the first studies to provide convolutional neural network and deep learning models to become popular in 2012. This network is developed by Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton [54]. There are basically consecutive layers of convolution and pooling. ReLU is used as activation function, and max-pooling is used in pooling layers. AlexNet uses 8-layer CNN. This network demonstrated for the first time that the features acquired through learning can surpass manually designed features and break the previous conventional patterns in computer vision. The general structure of AlexNet CNN model is shown in Figure 3.5.

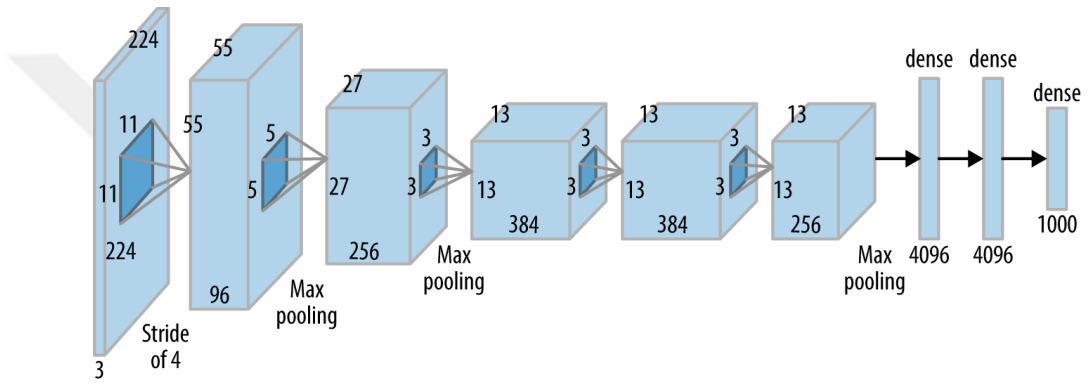


Figure 3.5 : The general structure of AlexNet architecture [8].

This deep convolutional neural network consists of 25 layers, 5 convolution layers, 3 max-pooling layers, 2 dropout layers, 3 fully connected layers, 7 ReLU layers, 2 normalization layers, softmax layer, input and classification (output) layers. In the first layer of AlexNet, the shape of the convolution window is 11×11 . A larger convolution window is required to understand the object. The convolution window shape on the second floor is reduced to 5×5 , then to 3×3 . In addition, after the first, second and fifth curvilinear layers, the web adds maximum deposition layers with a 3×3 window shape and a long pitch of 2. After the last convolutional layer there are two fully connected layers with 4096 outputs. These two large fully connected layers produce the model parameters approximately 1 GB of memory space. It is very important when considered the GPU memories.

In addition, AlexNet replaced the sigmoid activation function with a simpler ReLU activation function. On the one hand, the ReLU activation functionality is easier to

compute. On the other hand, ReLU activation functionality facilitates model training when using different parameter initialization methods.

3.4 VGGNet

VGGNet is a simple architecture network. It is developed by Simonyan and Zisserman in 2014 [55]. The VGG Network can be divided into two parts: The first part consists mostly of convolution and deposition layers and the second part consists of fully connected layers. The convolutional part of the network connects several VGG blocks in a row. A VGG block consists of a convolutional layer sequence followed by a maximum deposition layer for spatial down sampling. The original VGG network had 5 convolutional blocks, each of had two convolutional layer followed by three convolutional layers. The first block has 64 output channels, and each subsequent block doubles the number of output channels until it reaches 512.

There are two kinds of VGGNet which are commonly used. The different VGG models can be identified by the differences in the convolutional layer and the number of output channels in each block. The use of blocks results in very compact representations of the network definition and provides efficient design of complex networks. One kind of VGGNet is VGG16 which has 13 convolutions 3 fully connected layers. There are 41 layers including maxpooling, fully connected, ReLu, dropout and softmax layers in total. Approximately 138 million parameters are calculated. It has a structure in traditional sequential network architecture. A general structure of VGGNet is shown in Figure 3.6. The model training process is similar to that of AlexNet, except that it uses a slightly greater learning speed.

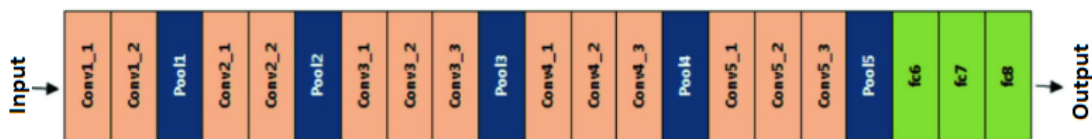


Figure 3.6 : The general structure of VGGNet architecture [9].

3.5 SqueezeNet

The SqueezeNet architecture was introduced in 2016 by Iandola et. al [56]. It was designed as a small neural network for the purpose of including fewer parameter, taking

up less space in memory and transmitting easily. It has an independent convolution layer (conv1), eight fire modules (fire2–9) and final convolution layer(conv10). It reduced the model size by decreasing the number of parameters 50 times. The graphical representation of SqueezeNet is given in Figure 3.7.

The difference from other networks is using fire block. Fire blocks consist of a 1×1 convolutional layer and it gives squeezed tensor as the output. After that, this tensor is fed to the 1×1 convolution layer and a 3×3 convolution layer and the outputs are concatenated. Another difference of this network is in its ability to handle images with different sizes as it has no fully connected layer.

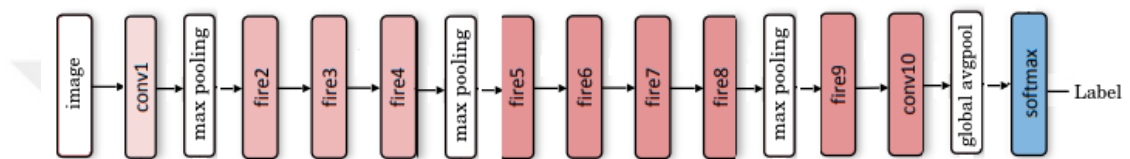


Figure 3.7 : The general structure of SqueezeNet architecture.



4. DATASET

There are few publicly available skin lesion datasets for the researchers. Table 4.1 shows these available datasets and including the number of images within each dataset. The two of the most popular datasets are PH^2 [57] and ISIC Challenge Dataset [58].

Table 4.1 : The datasets available for dermatological studies.

Datasets	# of Images
ISIC Archive [59]	33126
PH^2 [57]	200
DermoFit [60]	1300
Dermis [61]	397
MED-NODE [62]	170

4.1 ISIC- International Skin Imaging Collaboration Dataset

ISIC- International Skin Imaging Collaboration dataset has 13786 clinical cases in 2016 [58]. There are nine classes of skin lesions in that year. In 2020, the dataset has 33126 training and 10982 test images with two classes: benign or malignant melanoma [59]. Some sample images from this dataset can be seen in Figure 4.1.

It is publicly available from 2016 to 2020. The dataset has been provided as a challenge and it has been organized every year and the dataset has been updated each year. The studies using ISIC Archive dataset for lesion classification can be found in Table 4.2.

Table 4.2 : Some studies and their accuracy values for SLC over ISIC dataset.

Author of the studies	Accuracy(%)
Zhang et al. [63]	92.5
Aldwgeri et al. [64]	97
Albahar [27]	97.49
Sarkar et.al. [65]	99.50
Hosny et al. [45]	95.9
Hasan et al. [66]	93
Almaraz-Damian et al. [67]	92.4

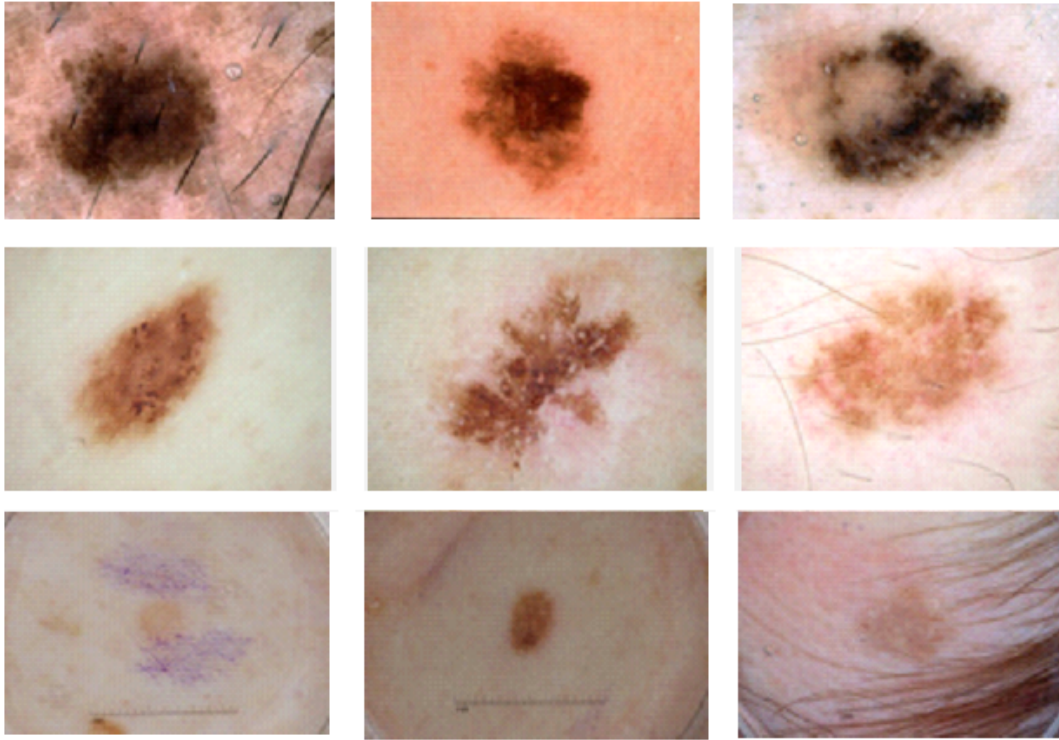


Figure 4.1 : Some sample images from ISIC dataset.

are provided to our experiments. The mostly used dataset is the ISIC dataset [68, 70] which also has ISBI challenge datasets [69] released in 2017, 2018, 2019, 2020, 2021 and 2022. The details of the dataset are

5. EFFECTS OF OBJECTS AND IMAGE QUALITY FOR SKIN LESION CLASSIFICATION

Malignant melanoma is one of the most rapidly increasing cancers in the world [11]. Despite its low prevalence (5%), it has higher mortality rate compared to other types of skin cancers. Just like many other types of cancer, early diagnosis of melanoma is very important for an effective treatment and to avoid poor prognosis. Computer aided diagnosis (CAD) systems can be used to assist physicians and to act as a second opinion.

Typically, images are collected using dermatoscopy, and lesions are segmented and classified as melanoma or not. The studies differ in their methods to process, segment and classify the lesions. Lesion segmentation is the primary step for melanoma detection. A typical dermatoscopy image has multiple lesions and suspected regions of interest. Each of these regions has to be processed differently using their characteristics such as border regularity, asymmetry and size of the lesion. The segmentation of the lesions can be performed via various classical segmentation algorithms. However, almost all recent studies employ deep neural network based models for lesion segmentation.

ISIC dataset which is commonly used in this problem has high quality images. However, the images taken by clinicians may not have such good quality due to imaging system, illumination and other issues in clinical environments. As a result, poor quality images with problems like brightness, blur, and noise can be acquired. Some of these problems cannot be alleviated using image processing techniques. In image recognition field, there are various studies about effects of image distortions on the output [71–76]. They analyzed the strengths and weaknesses of CNN models against image degradation. There are also a few studies about effects of image quality in skin lesion classification. One study investigate the CNN performance under different noise distortions in the skin images [77]. In another study, the researchers

used low quality images that are degraded with blurring, noise compression and channel errors to investigate the system performance [78].

In this study, we analyzed how the classification performance was affected by degradation and defects in lesion images. We considered contrast, noise and blur distortions and additional external objects such as hair and ruler. For this aim, we obtained new test sets and classify the images by using commonly used deep learning architectures for medical image analysis. We used four different CNN architectures which were commonly used for melanoma classification problem: AlexNet [79–81], VGG16 [80, 82, 83], ResNet50 [33, 34, 83], DenseNet [84, 85]. We investigated how these CNN models perform under image degradation and defects.

We performed our tests with many artificially produced data sets. These datasets are formed by changing image degradation level for blur, noise and contrast variations or by selecting images that contain objects such as hair or ruler in the ISIC data. Our purpose was to investigate the effects of image quality level and external objects on classification performance. In addition, we investigated the best CNN model that is effected the least from these degradation and objects.

5.1 Methodology

The proposed system has two main parts, training and testing. The block diagram of whole system and processes done in each part is shown in Figure5.1.

5.1.1 Dataset

In this study, we aimed to test our experiments on the mostly used dataset. As many other studies, we also used ISBI Challenge datasets that were released in 2017, 2018 and 2019 [3, 68–70]. In 2018 and 2019, the labels of training part of the dataset were given to the competitors, whereas the labels for the test and validation parts were not released. Hence, training, validation and test sets of 2017 and training sets of 2018 and 2019 are used for the experiments. We combined all images in these three datasets. After that, we split the dataset according to the class labels. There were labels for 9 and 7 classes in 2019 and 2018 datasets respectively. In the 2017 dataset, there were only seborrheic keratosis and melanoma classes.

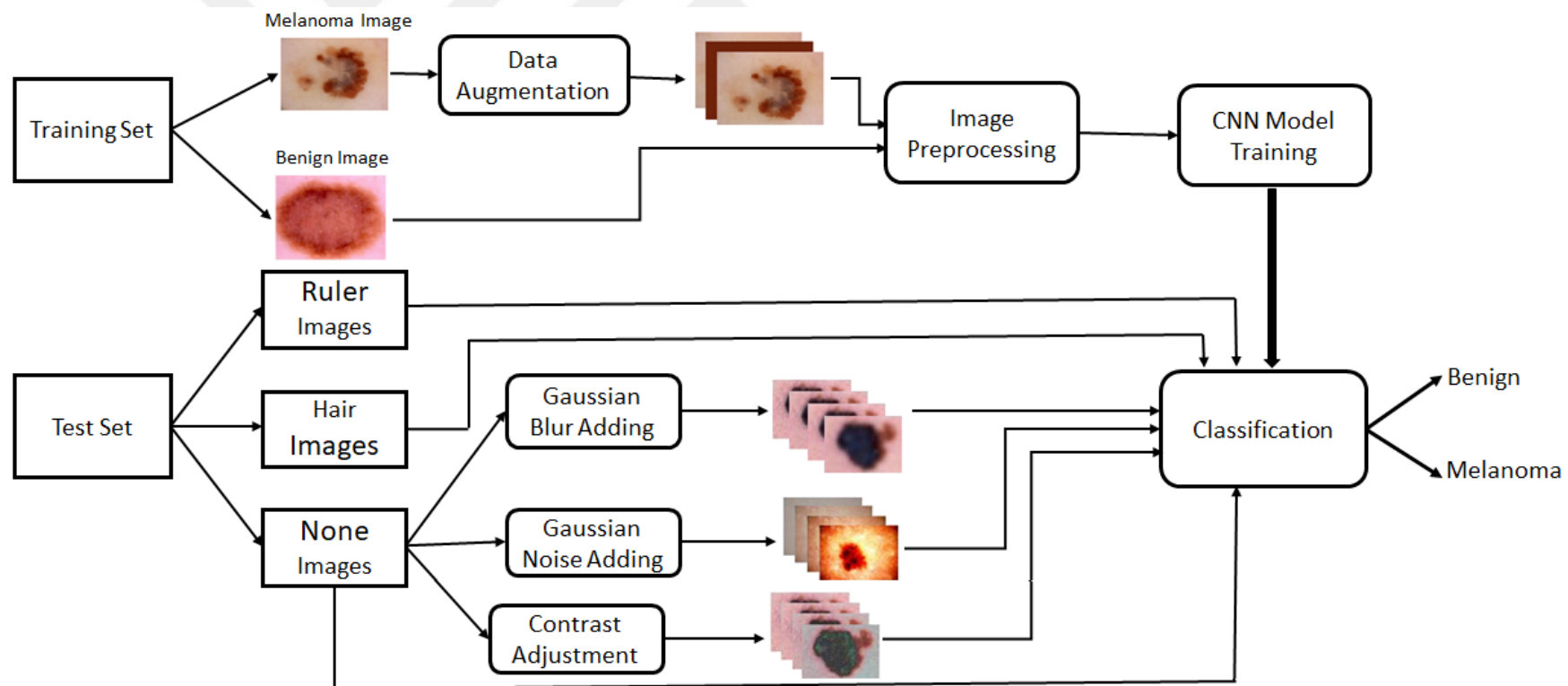


Figure 5.1 : Block diagram of whole system.

In this study, we dealt with only two classes benign and melanoma. Hence, lesions labeled as melanocytic nevus and benign keratosis and seborrheic keratosis classes were considered as benign. We divided the images in 2 classes and split the dataset into training and test partitions for both classes. We produced our partitions at approximately 80-20% ratio for training and test respectively.

In order to analyze the effects of image external objects, we formed 3 test subsets; $Test_{Ruler}$, $Test_{Hair}$, $Test_{None}$. However, training dataset had all images with and without hair and ruler. An artificial ruler was added to some of the images which are then added to $Test_{Ruler}$ set. For $Test_{Hair}$ dataset, we selected images that include any hair. It can be a few or more hairs. $Test_{None}$ set included only the lesion without hair or ruler. We discarded the images have both ruler and hair to understand the effects of the separately. The number of images per classes in the sets is shown in Table 5.1 and some sample images from the test sets are shown in Figure 5.2.

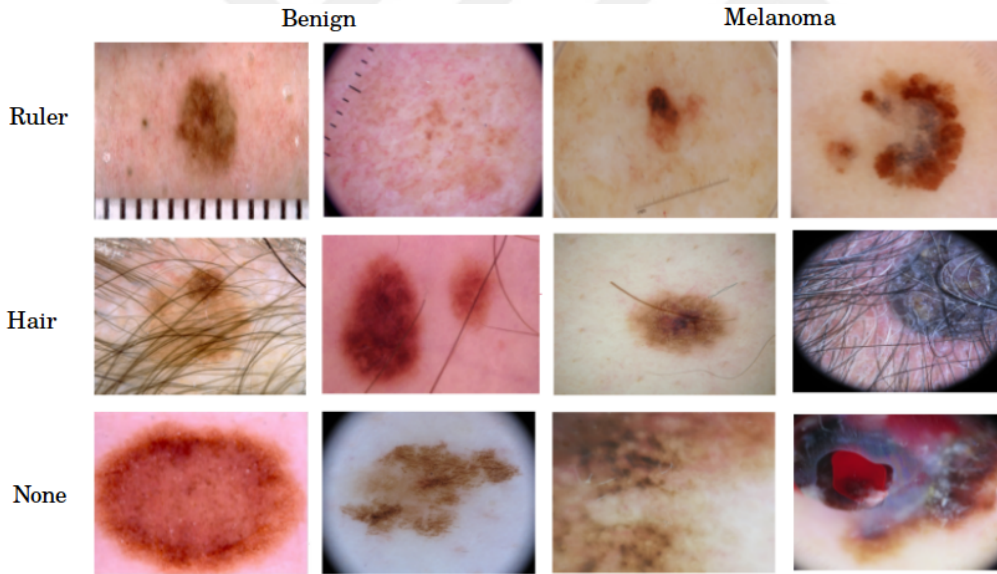


Figure 5.2 : Sample images from 3 test sets.

Table 5.1 : The number of images per classes in the sets.

Sets	# of Benign Lesions	# of Melanoma Lesions
Training	11400	3500
Test_Ruler	350	150
Test_Hair	1200	270
Test_None	1050	350

5.1.2 Data augmentation

There is imbalance in numbers of images in the classes as seen in Table 5.1. Therefore, we applied data augmentation processes on melanoma images in all sets to make the image numbers equal in the benign and melanoma classes . Rotation and flipping were used for augmentation. We rotated the images 45° , and flipped the images in horizontal, vertical and both axis. The augmented images are shown in Figure 5.3.

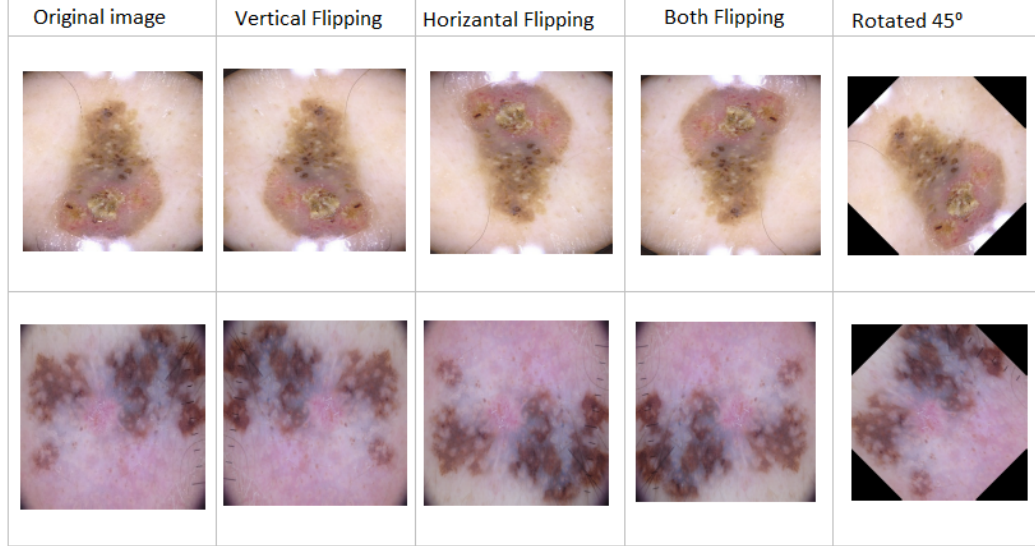


Figure 5.3 : Transformed images. The images in the first row are benign lesion, in the second row are melanoma lesions.

5.1.3 Test sets with degraded images

Image processing algorithms can have different performances under different conditions such as illumination, noise etc. We considered that Deep learning techniques, Convolutional Neural Networks can also be affected from these conditions. The quality of the images may adversely affect classification. Hence, we wanted to analyze the effects of using synthetically degraded images in melanoma classification. We used different levels of blur, noise and contrast degradation. We used the images in the $Test_{None}$ dataset for both classes.

5.1.3.1 Blurry image dataset

Blur is one of the main problems in image processing researches. It makes the classification harder, so many studies use deblurring algorithms as preprocessing. In this study, we did the opposite, we applied the Gaussian filters to our images. We change the filter sizes and obtain different blurring levels. For this aim, we changed

the value of σ from 1 to 20 and we calculated the filter size width and height according to the formula of $w, h = 2(2\sigma) + 1$. The filter size varied from (5 x 5) to (81 x 81) and we obtained 20 test sets of different levels of blur.

5.1.3.2 Noisy image dataset

Noise is inherent in all images with different levels. Hence, we added Gaussian distributed noise with zero mean and various scales (standard deviations) σ . We considered noise with scale (σ) varied from 1 to 30. Gaussian noise were added to all channels of skin lesion images separately. As a final step, image pixel values were normalized to 0- 255 range by MinMax Normalization algorithm. We obtained 30 different probe sets to analyze the effects of noise in melanoma classification.

5.1.3.3 Contrasted image dataset

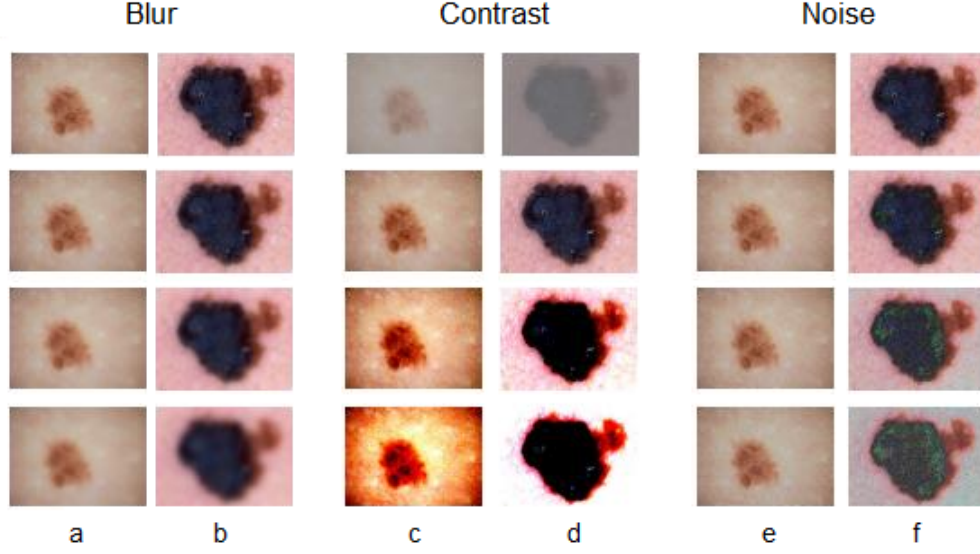
The images with different contrasts may also have different impact on classification performance. For this aim, we adjusted the contrast in this part of the study using a contrast factor. Contrast factor is a parameter where 0 means no contrast (all gray) and 1 means original contrast level. We obtained 20 different sets for the values of contrast factor from 0.2 to 4 increments of 0.2.

Some samples of artificially degraded images are shown in Figure 5.4. Also, number of images after data augmentation are given in Table 5.2.

Table 5.2 : The number of images per classes in the sets after data augmentation.

Sets	# of Benign Lesions	# of Melanoma Lesions
Training	11400	14000
Test_Ruler	350	300
Test_Hair	1200	1350
Test_None	1050	1050
Test_Blurred	1050	1050
Test_Noisy	1050	1050
Test_Contrasted	1050	1050

After specifying training and test sets, we applied pre-processing steps to our images. Before training the networks, we first resized the images (224×224) as expected input size of the CNN Models. We also normalized the images with a predefined function in Pytorch library [86]. This function normalizes each channel (RGB) of the image with



a,c,e columns are benign; b,d,f columns are melanoma images.
 For Blur set, they are blurred by 5x5, 21x21, 41x41, 81x81 filters,
 For Contrast set, they are enhanced contrast by factors of 0.2, 1, 2, 4,
 For Noise set, the noise with sigma =1, 10, 20, 30 added respectively.

Figure 5.4 : Sample images from synthetic image test sets.

the following formula

$$x(c, i) = \frac{x(c, i) - \mu_c}{\sigma_c}$$

where $x(c, i)$ is the image pixel i in channel c ($c \in \{R, G, B\}$), μ_c and σ_c are the mean and scale of channel c respectively. μ_c and σ_c are computed using all images in ISIC datasets.

5.2 Experimental Results

In this study, we removed the last fully connected layer of all networks and added our custom classifier. This classifier had a fully-connected network, which had 1000 input neurons and a custom hidden layer which has 40 input neurons, also the output neuron number was set to 2 due to the binary nature of classification. We also defined the activation function as ReLU and a dropout was set to 0.2. After that, we defined the loss function and optimizer in our network. We used NLL (Negative log-likelihood) Loss Function and ADAM optimizer for this problem. We also set the learning rate as 0.0001. After all definition, the model was prepared for training.

We trained our models until training loss is lower than 10^{-6} . It is the early stopping technique to avoid overfitting. After training, we saved the models and test sets are tested with same models for each CNN architecture.

The success of the model is related to the number of samples assigned to the correct class and the number of samples to the wrong class. The performance metrics obtained as a result of the tests can be expressed with the confusion matrix. In the confusion matrix, the rows represent the actual numbers of the samples in the test set, and the columns represent the model's prediction. The confusion matrix is shown in Table 5.3.

Table 5.3 : Confusion matrix.

	Classified as Melanoma	Classified as Benign
Diagnosed as Melanoma	True Positive (TP)	False Negative(FN)
Diagnosed as Benign	False Positive(FP)	True Negative (TN)

The most popular and simple method used to measure model performance is the accuracy rate of the model. It is the ratio of the number of correctly classified samples (TP + TN) to the total number of samples (TP+TN+FP+FN). Error rate is the ratio of the number of misclassified samples (FP+FN) to the total number of samples (TP+TN+FP+FN).

We evaluated the results of the trained models with different metrics which were commonly used in classification problems. These metrics are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.1)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (5.2)$$

$$Specificity = \frac{TN}{FP + TN} \quad (5.3)$$

$$Precision = \frac{TP}{TP + FP} \quad (5.4)$$

where TP and TN mean true positive and true negative and FP and FN mean false positive and false negative respectively.

We divided our test sets into overlapping 10 folds. All metrics were calculated for each fold and they are averaged over ten folds. Figure 5.5, 5.6 and 5.7 show that the performance metric results that were calculated over the noisy, blurry and contrasted image datasets, respectively.

In Table 5.4, 5.5 and 5.6, we calculated the accuracy, sensitivity, specificity, precision and true_predicted-all metrics over other three test sets: hair, none and ruler respectively by using 4 different CNN models. Accuracies over three test sets are shown in Figure 5.8.

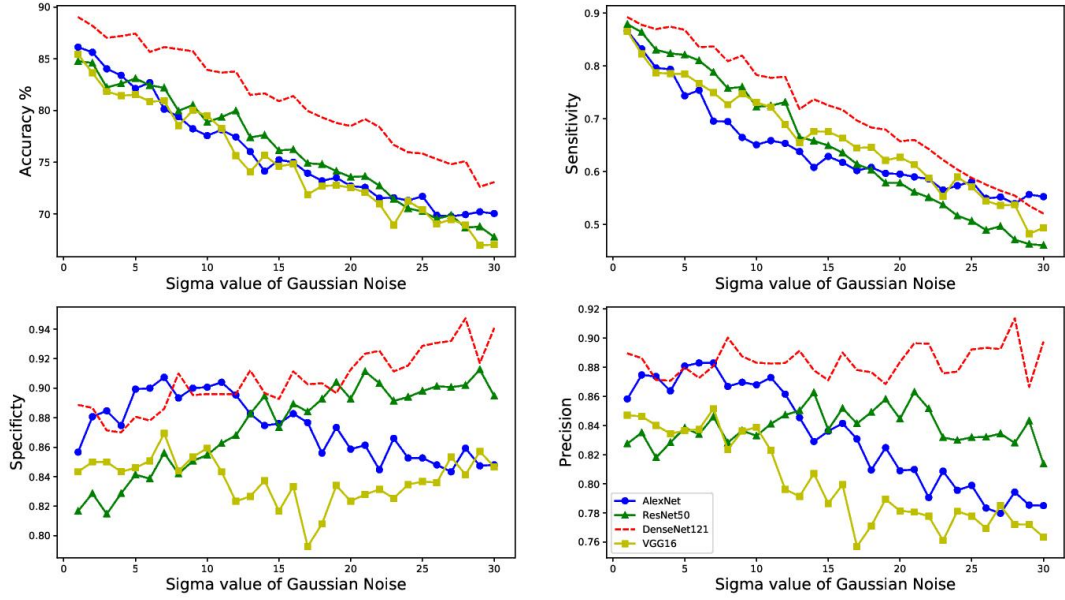


Figure 5.5 : Graphics of metrics over noisy images.

Table 5.4 : Performance metrics of the models for *Test_Hair* dataset (TP: True-Predicted).

CNN Models	Accuracy	Sensitivity	Specificity	Precision	TP/All
DenseNet121	89.22	0.8955	0.8883	0.9002	2275/2550
ResNet50	87.96	0.8681	0.8925	0.9008	2243/2550
AlexNet	84.51	0.8688	0.9133	0.8676	2155/2550
VGG16	86.2	0.8177	0.9116	0.9123	2198/2550

It is crucially important to predict diseases correctly. Therefore, we measured True Negative Ratio (TNR) on benign lesions and True Positive Ratio (TPR) on melanoma

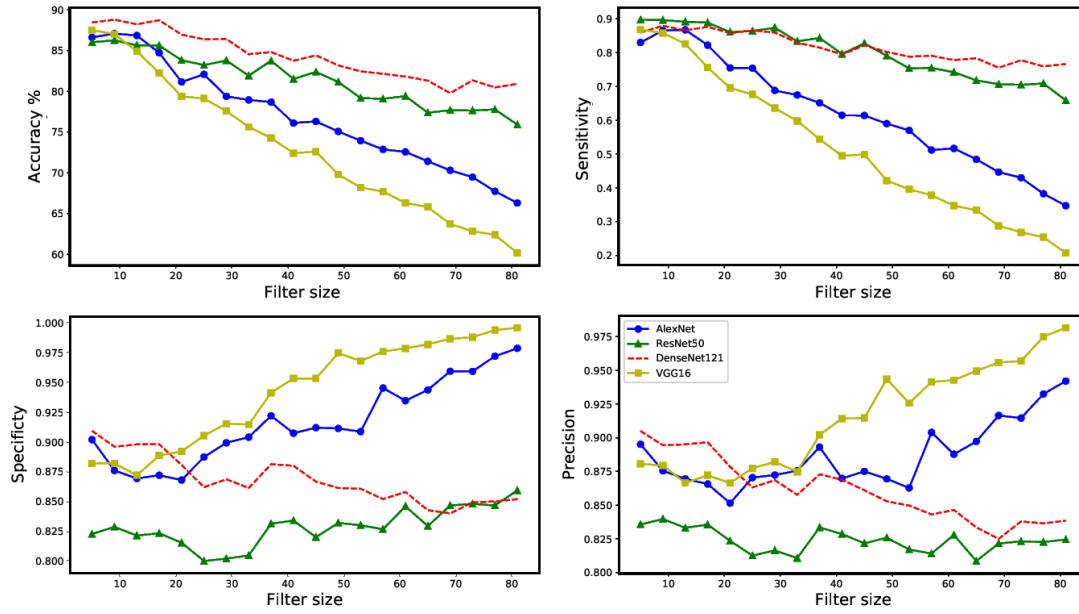


Figure 5.6 : Graphics of metrics over blurry images.

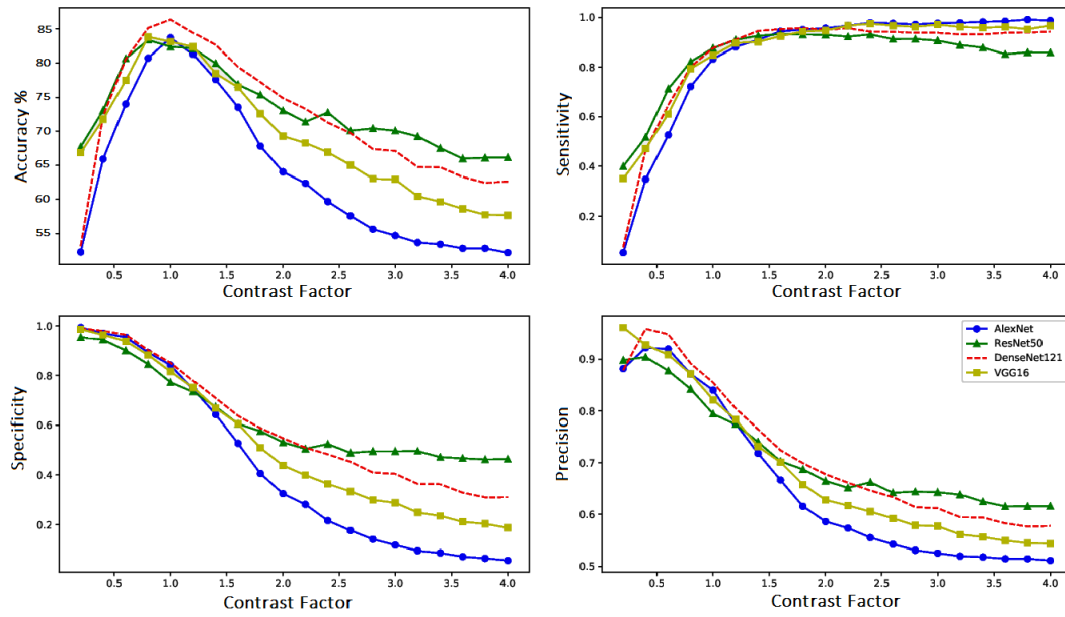


Figure 5.7 : Graphics of metrics over contrasted images

Table 5.5 : Performance metrics of the models for *Test_None* dataset (TP: True-Predicted).

CNN Models	Accuracy	Sensitivity	Specificity	Precision	TP/All
DenseNet121	88.81	0.9209	0.8552	0.8641	1865/2100
ResNet50	86.95	0.9142	0.8247	0.8391	1826/2100
AlexNet	84.95	0.8657	0.8333	0.8391	1784/2100
VGG16	86.57	0.8161	0.9152	0.9059	1818/2100

Table 5.6 : Performance metrics of the models for *Test_Ruler* dataset (TP: True-Predicted).

CNN Models	Accuracy	Sensitivity	Specificity	Precision	TP/All
DenseNet121	86	0.9100	0.8171	0.8100	559/650
ResNet50	85.23	0.8933	0.8171	0.8072	554/650
AlexNet	80.92	0.8466	0.7771	0.7650	526/650
VGG16	81.69	0.9433	0.7085	0.7350	531/650

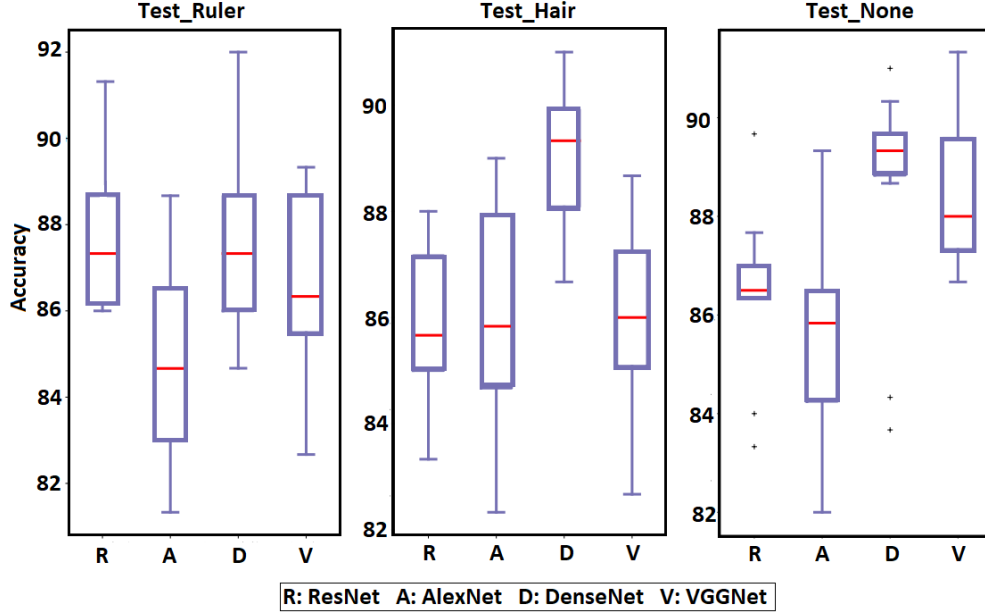


Figure 5.8 : Accuracies of deep neural network models for test sets that are calculated over 3 datasets.

images separately for four models. It is given in Table 5.7.

$$TPR = \frac{TP}{TP + FN} \quad (5.5)$$

$$TNR = \frac{TN}{FP + TN} \quad (5.6)$$

In this study, we also dealt with some other issues and the most important one is overfitting problem in training of the networks. We performed many strategies to eliminate overfitting of data. Our strategies that we used were as follows:

- Loss functions: We tried two different loss functions: Negative Log Likelihood loss function and Cross Entropy Loss Functions.

Table 5.7 : Accuracy of the models for all test sets for both classes.(R: ResNet, A:AlexNet, D: DenseNet, V: VGG16)

Test Sets	Benign TNR				Melanoma TPR			
	R	A	D	V	R	A	D	V
Ruler Set	87.73	77.6	83.33	83.6	87.73	92	91.73	89.86
None Set	90.73	84.8	90.86	91.53	82.13	86.2	86.2	85.33
Hair Set	93.66	88.13	93	89.53	78.06	83.73	85.06	82.46
Blur Set (9x9)	87.6	82.86	89.6	88.2	86.53	89.6	87.93	85.8
Blur Set (77x77)	97.2	84.66	85	99.4	38.26	70.86	75.93	25.4
Noise Set ($\sigma = 2$)	88.06	82.86	88.66	85	83.2	86.33	87.73	82.26
Noise Set ($\sigma = 28$)	85.93	90.2	94.73	84.13	53.93	47.13	55.46	53.73
Contrast Set ($c = 0.4$)	97	94.4	97.93	96.33	34.8	51.73	46.53	47.13
Contrast Set ($c = 3.8$)	6.53	46.26	31	20.33	99.06	85.93	93.86	95.2

- Optimizer: ADAM and Stochastic Gradient Descent (SGD) optimizers were handled for the experiments.
- Weight decay: The weight decay parameter was set from 0.1 to 0.9, and we did our experiments with 9 different values of this parameter.
- Early stopping: The training loss was calculated after each epoch and if the loss was lower than a predefined value, the training was cut at that epoch.
- Learning rate: Different values of learning rate parameter were set for the training.

After trying all strategies, we obtained the minimum training loss with early stopping technique. We set the expected loss as $1e^{-6}$. If the calculated loss is lower than $1e^{-6}$, we stopped the training at that point.

6. ENSEMBLE OF CONVOLUTIONAL NEURAL NETWORKS FOR SKIN LESION CLASSIFICATION

Although nearly everybody has benign melanocytic nevi on the skin, their malignant counterpart, melanoma, is relatively rare, but it causes approximately 75% of deaths among skin cancers. It is the 5th common cancer with around 0.1 million annual new cases in USA [11]. Although melanoma is not included among the most common types of cancer in Turkey, nearly two thousand new cases were reported and 888 of them resulted in death in 2020 [4].

Melanoma incidence has doubled in the last 30 years both in Turkey and all over the world. It is usually diagnosed by clinicians, mostly dermatologists, on physical examination. Nevertheless, to differentiate benign melanocytic nevi and melanoma can be a difficult task for clinicians. Some cases cannot be diagnosed without histopathologic examination.

Early diagnosis plays a very important role in all types of cancer. With early diagnosis, the progression of the disease can be brought under control, and with the necessary interventions it can be cured.

Early detection of melanomas is crucial to achieve good survival rates. However, especially in the early stages of melanoma, its differentiation from benign nevi is a challenge for clinicians. For this aim, the researchers are willing to help to the clinicians and many Computer Aided Diagnosis (CAD) systems have been proposed to detect melanoma in early stages.

In this study, we proposed a CAD system for classification of the skin lesions as melanoma or benign melanocytic nevus. We used the ISIC dataset which is prepared for 2020 challenge. We applied the dataset on 11 different Convolutional Neural Network (CNN) models and analyzed the test results.

6.1 Methodology

In this study, the proposed approach consists of fusing of five different convolutional neural networks. The system architecture is given in Figure 6.1.

6.1.1 Preprocessing

The amount of data used to train CNN has very important role on its performance. In this study, we used 565 melanoma and 600 benign nevi images. We divided our dataset into training and test clusters with the rate of approximately 80% and 20% respectively. After that, we split the training set into 5 folds, so in each fold there are 96 benign and 90 melanoma images. Due to the limited number of images in folds, the data augmentation process is a necessity for classification of skin lesions by using deep learning algorithms. Hence, we applied preprocessing to images in training set in order to increase the number of data. For this aim, we applied rotation and flipping operations. We rotated the images by increasing degree of 15 between 0-360° and flipped the image in vertical, horizontal and both directions. With this augmentation, we increased training data by 27 times.

6.1.2 Training and ensemble of the models

Pre-trained CNN models are commonly used in skin lesion classification problems. Mostly used CNN models are ResNet applied in [29, 31, 87–89], AlexNet utilized in [31, 90, 91], VGGNet used in [31, 91, 92], DenseNet handled in [87, 89, 93] and Inception Net used in [87, 89, 94]. Therefore, we decided to use pretrained CNN models in our CAD system. We selected 5 different convolutional neural network models and used their different variations of those models, so we handled 11 CNN models to train. These models are: DenseNet121, DenseNet161, DenseNet169, DenseNet201, ResNet18, ResNet50, VGGNet19, VGGNet16_bn, SqueezeNet1_1, SqueezeNet1_0, AlexNet.

We changed the last layer of those pretrained models and we added two extra fully connected layers instead of the last fully connected layer. New fully connected layers had 1000 input and 40 output neurons and 40 input and 2 output neurons, respectively.

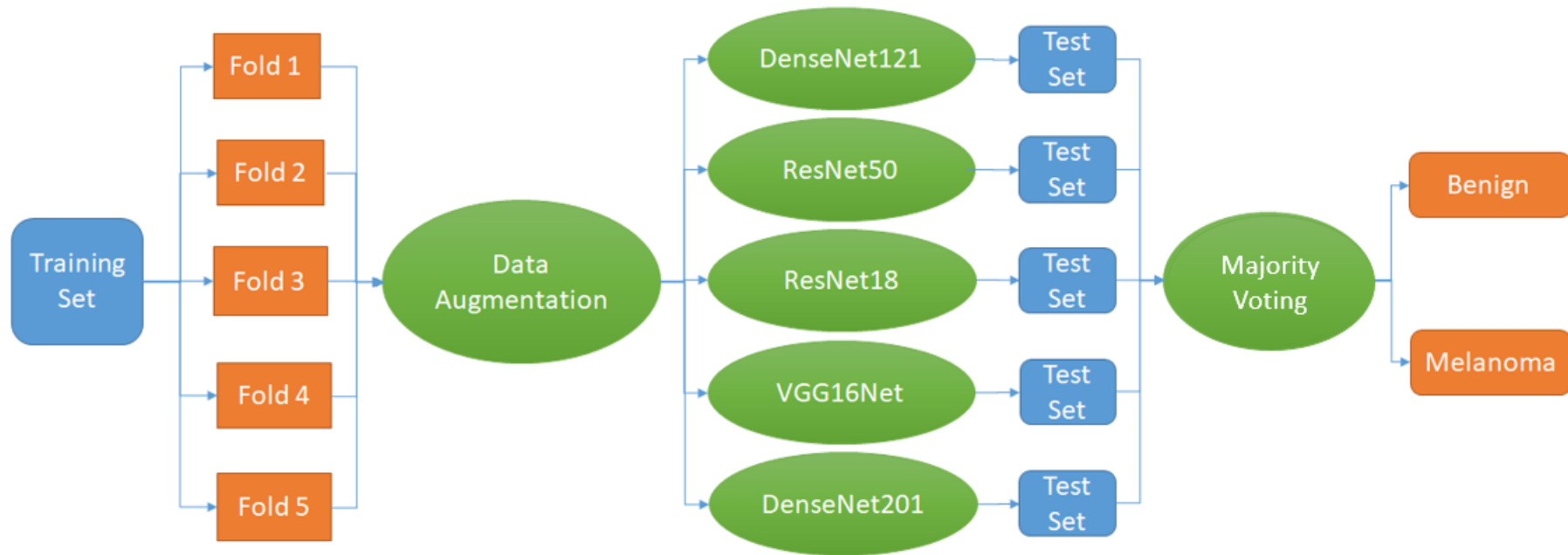


Figure 6.1 : The overall system architecture.

We used same parameter values for each model in training and the parameters in the models are given in Table 6.1.

Table 6.1 : Parameter values in the models.

Parameter	Value
Number of Epochs	100
Learning rate	0.0001
Batch size	32
Dropout	0.8
# of neurons in fc layer(input)	1000
# of neurons in fc layer(input)	40
# of neurons in fc layer(output)	2

We trained the models with maximum 100 epochs. However, we used early stopping criteria. If the training loss is less than 10^{-7} and the training accuracy is greater than 99%, then we stopped the training of the model and saved the model. On the other hand, if the criteria were not satisfied, then the model was trained until the epoch number reaches 100.

The usage of hybrid or ensemble Deep Neural Networks has become quite common in classification problems [44, 64, 80, 95, 96]. We also tried to use an ensemble of CNN models in which we trained each fold with a different CNN model. We specified the 5 different models for 5 folds and it is given in Table 6.2.

Table 6.2 : Specified CNN models for every fold.

Fold number	CNN Model
Fold 1	ResNet18
Fold 2	VGG19Net
Fold 3	DenseNet201
Fold 4	ResNet50
Fold 5	DenseNet169

6.1.2.1 Majority voting algorithm (MV)

The CNN models are trained over each fold and the outputs of the each model are ensembled by voting strategy. Simple majority voting algorithm is used to decide the final class labels of test subjects [97]. The classifiers (voters) give a label to each test instance. It compares the class labels of all voters and if the majority (more than half

of the voters) is achieved per each test image final decision is specified by majority voting. It is used for ensemble of the models in skin lesion classification. Hence, we wanted to provide a more accurate decision mechanism using the this voting algorithm.

6.1.2.2 Averaging probabilities algorithm (AP)

CNN model gives the probability values of being in a class and according to this probability value, it assign the input to a class. We took this value instead of using the class label directly, and averaged five values of five different models and based on the calculated new probability value in this method. After that, we compared the new probability value with a threshold and assigned the class label to that image. We used this approach to ensemble the five CNN models outputs. We called this approach as averaging probabilities (AP).

6.1.2.3 Weighting with confidence ratio algorithm (WCR)

In this approach, first step is to calculate the confidence ratio. We calculated the accuracy performance of each model on each fold. The measured performance value for each fold determines our confidence in the results produced in that model. Hence, we multiplied the probability values and the confidence ratio of each model and we obtained a weighted values. We averaged these results and after thresholding, new class labels were assigned.

6.2 Experimental Results

In our experiments, we obtained 5022 images with data augmentation operations for both malignant melanoma and benign nevi classes in five folds of training. We trained each fold with a different CNN model and we saved them.

There are 235 images in the test set. It includes 120 benign nevi and 115 melanoma samples. We tested our system with different 11 saved models over test set and the evaluation metrics of each model are shown below. Figure 6.2, Figure 6.3 and Figure 6.4 show the metrics are measured by using maximal voting, averaging probabilities and weighting with confidence ratio algorithms.

In addition, we compared performances of ensemble algorithms. The comparison of accuracies of the three algorithms is shown in Figure 6.5.

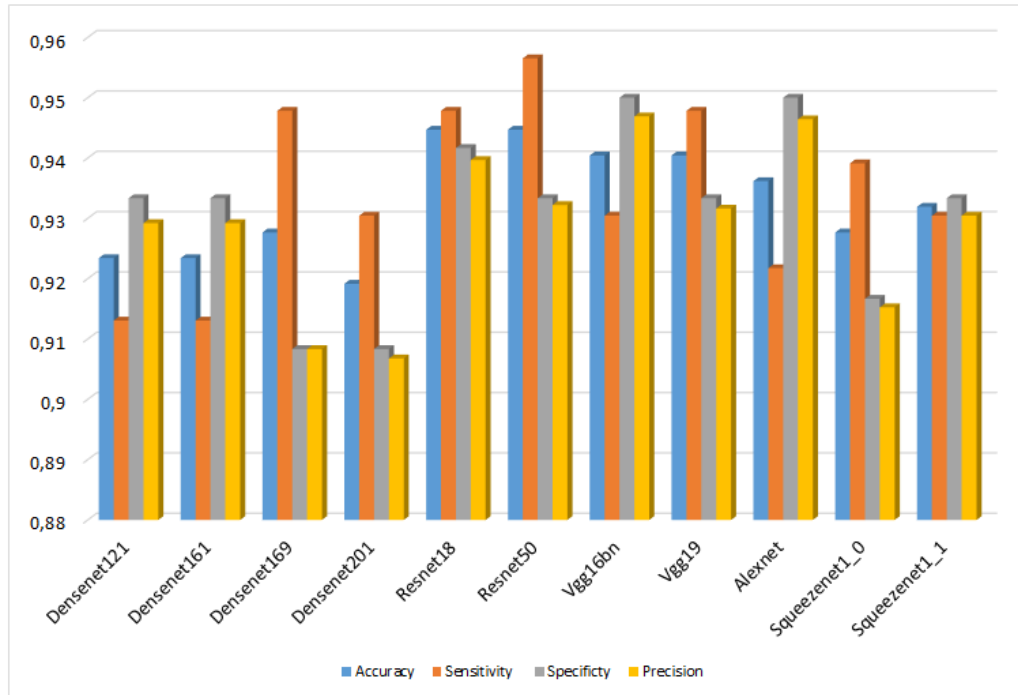


Figure 6.2 : The statistics for each model on test set by using majority voting approach.

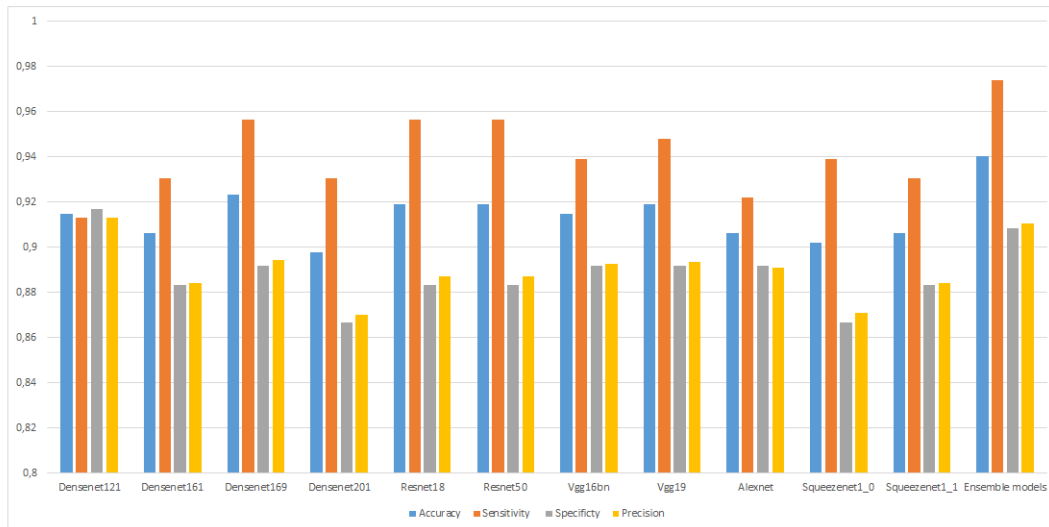


Figure 6.3 : The statistics for each model on test set by using averaging probabilities approach.

The model statistics of each model on each fold and their ensembled results is shown in Figure 6.6 and Table 6.3.

Furthermore, Table 6.4 lists the number of images which are predicted correctly.

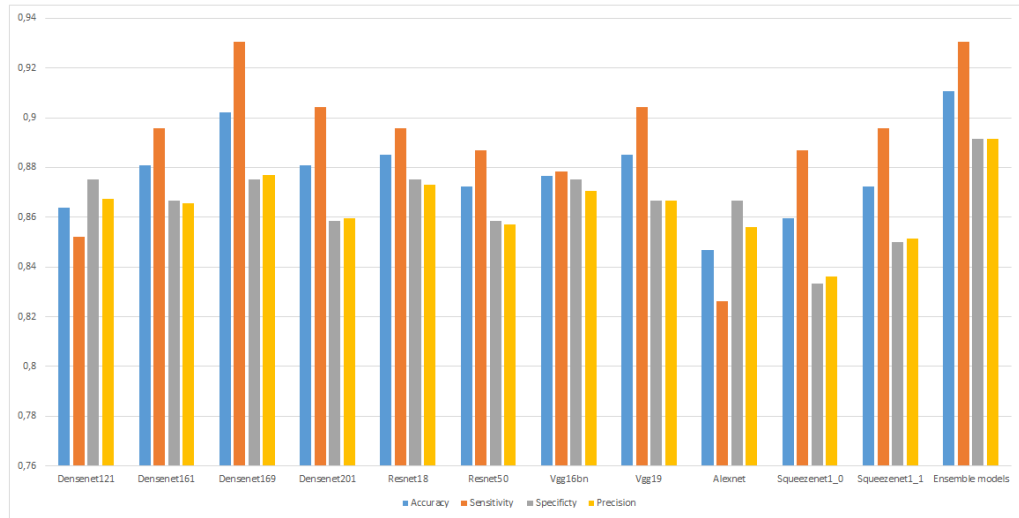


Figure 6.4 : The statistics for each model on test set by using weighting with confidence ratio approach.

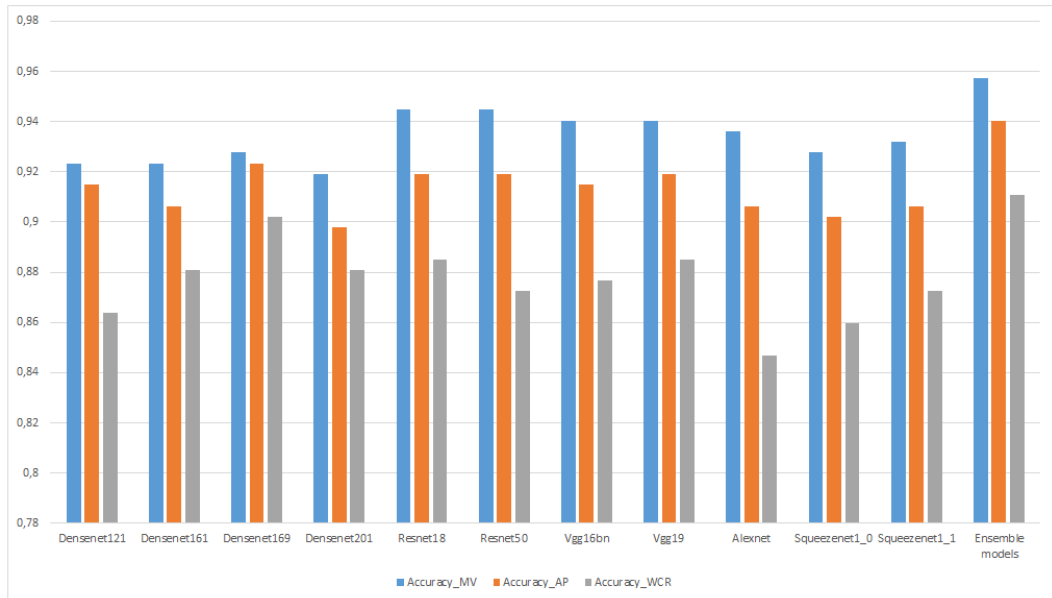


Figure 6.5 : The comparison of the accuracies of three ensemble algorithms .

We also analysed the three different ensemble techniques and their results. Majority voting is better than the other approaches in terms of positive prediction rates. The results of three approaches per models are shown in Table 6.5.

The best performances are obtained with ensemble of models. The proposed approach predicted incorrectly for only 10 images in our test set and these images are shown in Figure 6.7.

We achieved the classification rate of benign nevi as 93.3% and melanoma as 98%. The overall accuracy of our proposed model was calculated as 95.76%. We compared the

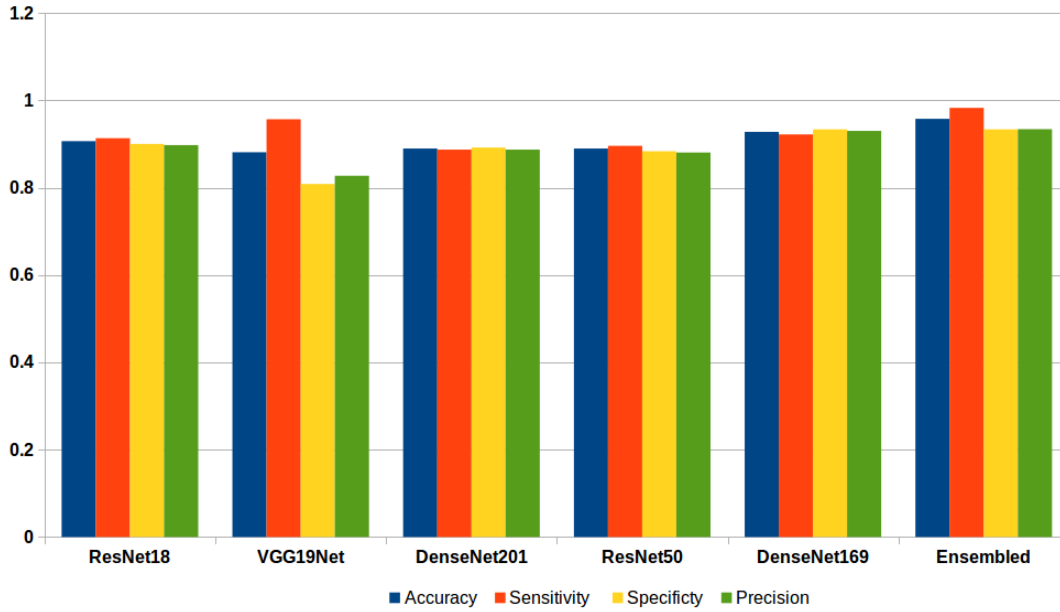


Figure 6.6 : The model statistics of each model on each fold and ensemble of all models.

Table 6.3 : The model statistics of each model on each fold and ensemble of all models.

Models	Accuracy	Sensitivity	Specificity	Precision
ResNet18	0.906	0.913	0.9	0.897
VGG19Net	0.880	0.956	0.808	0.827
DenseNet201	0.889	0.886	0.891	0.886
ResNet50	0.889	0.895	0.883	0.880
DenseNet169	0.927	0.921	0.933	0.929
Ensembled	0.957	0.982	0.933	0.933

Table 6.4 : True predicted ratios of models.

Models	Benign_TNR	Melanoma_TPR
ResNet18	0.9	0.913
VGG19Net	0.808	0.95
DenseNet201	0.891	0.886
ResNet50	0.883	0.895
DenseNet169	0.933	0.921
Ensembled	0.933	0.982

proposed system performances with other previous studies used ISIC dataset for skin lesion classification problem in 2020 and 2021. A comprehensive list of skin cancer detection systems using CNN classifiers is presented in Table 6.6.

Table 6.5 : Performance metrics for all models. (MV:Majority voting, AP:Averaging probabilities, WCR: Weighting with confidence ratio.)

Model	Approach	Accuracy	Sensitivity	Specificty	Precision
Densenet121	MV	0,92	0,91	0,93	0,93
	AP	0,91	0,91	0,92	0,91
	WCR	0,86	0,85	0,88	0,87
Densenet161	MV	0,92	0,91	0,93	0,93
	AP	0,91	0,93	0,88	0,88
	WCR	0,88	0,90	0,87	0,87
Densenet169	MV	0,93	0,95	0,91	0,91
	AP	0,92	0,96	0,89	0,89
	WCR	0,90	0,93	0,88	0,88
Densenet201	MV	0,92	0,93	0,91	0,91
	AP	0,90	0,93	0,87	0,87
	WCR	0,88	0,90	0,86	0,86
Resnet18	MV	0,94	0,95	0,94	0,94
	AP	0,92	0,96	0,88	0,89
	WCR	0,89	0,90	0,88	0,87
Resnet50	MV	0,94	0,96	0,93	0,93
	AP	0,92	0,96	0,88	0,89
	WCR	0,87	0,89	0,86	0,86
Vgg16bn	MV	0,94	0,93	0,95	0,95
	AP	0,91	0,94	0,89	0,89
	WCR	0,88	0,88	0,88	0,87
Vgg19	MV	0,94	0,95	0,93	0,93
	AP	0,92	0,95	0,89	0,89
	WCR	0,89	0,90	0,87	0,87
Alexnet	MV	0,94	0,92	0,95	0,95
	AP	0,91	0,92	0,89	0,89
	WCR	0,85	0,83	0,87	0,86
Squeezenet1_0	MV	0,93	0,94	0,92	0,92
	AP	0,90	0,94	0,87	0,87
	WCR	0,86	0,89	0,83	0,84
Squeezenet1_1	MV	0,93	0,93	0,93	0,93
	AP	0,91	0,93	0,88	0,88
	WCR	0,87	0,90	0,85	0,85
Ensembled	MV	0,96	0,98	0,93	0,93
	AP	0,94	0,97	0,91	0,91
	WCR	0,91	0,93	0,89	0,89

As seen from the Table 6.6, our CAD system have promising results. Therefore, we believe the proposed ensemble classifier has competitive performance compared to similar systems given in literature.

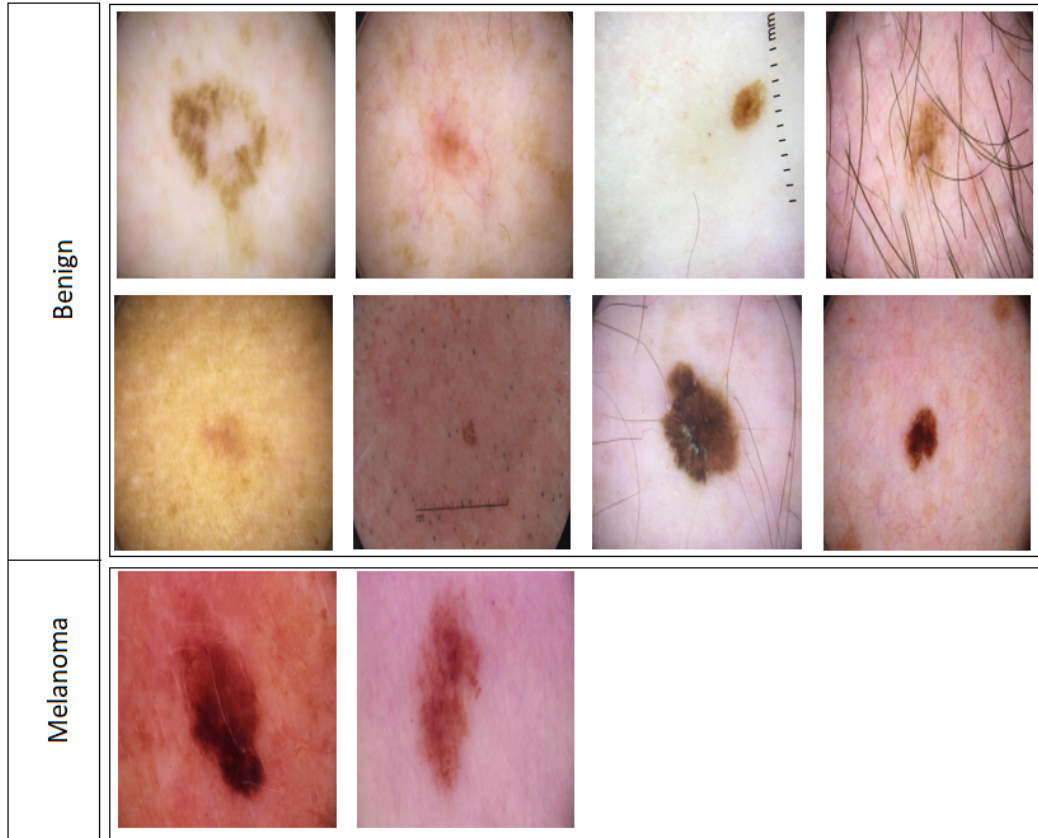


Figure 6.7 : The images classified wrongly.

Table 6.6 : A comprehensive list of skin cancer detection systems using CNN classifiers.

Author	Year	Accuracy (%)
Hasan et.al. [98]	2021	97
Ours	2021	96
Wei et.al. [99]	2020	96
Harangi et.al. [100]	2020	95
Kassem et.al. [101]	2020	94
Sagar et.al. [87]	2020	93
Almaraz et.al. [102]	2020	92
Tan et.al. [103]	2020	88

7. DISCUSSION

In this study, we focused on melanoma recognition system using convolutional neural networks from two different points of view. The findings in each study are discussed, and suggestions for future work are presented in this section.

There are some limitations common to our both studies. The first limitation was the size of dataset. Especially the number of images containing melanoma lesions were small. In order to deal with the unbalanced dataset, we tried to equalize the numbers of the images in the classes by reducing the benign lesion images. Hence, we obtained limited size dataset. This may cause overfitting issues and limited generalization in model training.

Furthermore, occlusion is another limitation of both studies. We obtained the images from a commonly used dataset (ISIC). There are many images with hair and/or rulers near or on the lesions. If most of the lesion area are occluded by hair or ruler, our system may fail to recognize the lesion.

Additionally, using a dataset that is composed of images taken under uniform conditions may not be useful to train a CNN model that would work at clinic environment. It would be better to have images taken at different imaging conditions (from multiple health centers) with different quality dermatoscopes.

Besides these common limitations, there are also specific limitation to each study, which are discussed in next subsections.

7.1 Discussion about Effects of Objects and Image Quality for Skin Lesion

We investigated the melanoma classification performance with different CNN models with image degradation and distortion problems. Four different CNN models were selected and six different types of degradations were considered.

The best performing model was DenseNet121 against all models in almost all situations. The all layers in DenseNet121 are densely connected to each other and all

information is transferred to all layers. Hence it allows the feature reuse. In addition, it includes 121 layers and it is deeper than the other models considered in the study. Due to these features of DenseNet121, we can interpret that it produced the best results in the problems we dealt according to our observations.

The effect of additional objects such as hair and rulers on classification performance was investigated in the study. Images with ink stains also exist in the dataset. The ink stains can be considered as another extra object in lesion images and analyzed within the same framework for further studies. Moreover, resolution may be another issue that can be addressed in this study. Classification of melanoma on images with different resolutions is one of the planned works.

7.2 Discussion about Ensemble of Convolutional Neural Networks for Skin Lesion Classification

A convolution neural network based computer aided melanoma recognition system was proposed in this study. To achieve higher accuracy, we used ensemble learning. We ensembled five CNN models in training and testing. Their results were fused with majority voting technique. During testing process, an image can be labeled as benign by one CNN model and melanoma by another. In such a case, final label is selected according to response of majority. The ensembled model has achieved better accuracy than a single CNN model. We used 5 different models as they could learn different features of the lesions. By using these models together, we achieved improved classification performance.

In majority voting method, only label outputs are taken into account. Probability values assigned by classifiers are not considered. On the other hand, in averaging probabilities (AP) and weighting with confidence ratio (WCR) methods consider only the probability values instead of label. Hence, averaging of probability values decrease the final probability value if one classifier gives a different label. Hence, we observed that AP and WCR methods had lower performance.

A limitation of this study is using the images directly without any preprocessing such as image cleaning to eliminate irrelevant information that may degrade classification performance. For this aim, segmentation procedure can be involved to the system, and

classification can be performed only over the segmented lesions. Thus, background can be eliminated with any unnecessary external objects.

In addition, different strategies can be used in the system. We can utilize generative adversarial networks to generate synthetic images with same statistics of the original images in the training set [104]. This strategy can overcome the limitation caused by the small number of images.

In the proposed system, only images are used for classification. However, there are other forms of clinical data such as demographics, skin type etc which can be fed into network to improve classification performance. It would be a useful extension to our system.

The CAD system classified some benign lesions as melanoma. Therefore, we can work with the dermatologists to reanalyze the misclassified lesions by checking if there are any visible differences or by giving information about key features between the benign and melanoma lesions. It can help us to resolve and decrease the number of misclassifications.

The increasing complexity of models used for artificial intelligence require more transparency and interpretability in explaining how decisions are made. It is thought that the explainable AI (XAI) field can help in ensuring transparency while taking advantage of the efficiencies provided by artificial intelligence. Explainable AI refers to the methods and techniques of applying artificial intelligence technology (AI) in such a way that the results of the solution can be understood by human experts. XAI techniques are recently used in applications of deep learning and machine learning [105,106]. XAI techniques used in medical image processing can be categorized as visual, textual, and example-based methods [107]. Visual explanation namely saliency mapping is the most commonly used form of them. It consists of three sub approaches: back propagation-based, perturbation-based or multiple instance learning-based approaches. These approaches are also divided into some other approaches and one approach is class activation mapping (CAM) used in back propagation-based approaches in XAI techniques [108]. It was specified that CAMs could be handled in ensembles of CNNs as interpretable technique [109]. We used pretrained CNN models based on the back propagation method in this study. However,

we did not include any of these explanation approaches so that it is another limitation of our study. To overcome this limitation, we are planning to provide XAI by adapting one of these approaches, especially CAMs, to our CAD system.



8. CONCLUSION

In this thesis, skin lesion classification problem is investigated with two different studies. In the first study, we used a public dataset and investigated the effects of external objects and image quality issues on different deep learning models. In the second study, we proposed a skin lesion classification system that works on ISIC dataset.

8.1 Conclusion of Effects of Objects and Image Quality for Skin Lesion

In this study, skin lesion classification is investigated from a different perspective that is typically not covered in other studies. We investigated how the classification performance is affected when image quality is poor, and some parts of the lesion are occluded with external objects such as hair or ruler.

We wanted to test the feasibility of melanoma detection from images that are taken with mobile phones or non-professional cameras under poor imaging conditions. We investigated the level of distortions up to which melanoma detection can be achieved at clinically acceptable performance. In addition, we investigated the performance of commonly used architectures when there are external objects in the image.

We investigated 4 different CNN models that are frequently used and compared their performance with image quality issues and external objects. We wanted to demonstrate which CNN model is more robust to image degradation and create a guide for future studies.

We used the datasets from ISIC Challenges, and we split the dataset for training and test purposes. We formed 3 subsets for testing that are ruler (Test_Ruler), hair (Test_Hair) and lesion (Test_None: images with lesion only). Then, we applied Gaussian kernels to obtained blurred images by using 20 different σ values, added zero mean Gaussian noise with 30 different standard deviation (σ) values. In addition, we modified the contrast of the images with a contrast factor that varies from 0.2 to 4.0.

We used four pretrained CNN architectures; AlexNet, DenseNet121, ResNet50 and VGG16. We replaced the last FC layer with our classifier. It consisted a FC layer with one input layer with 1000 neurons, one hidden layer with 40 neurons and one output layer with 2 neurons. After the training with four different models, we tested the system with 6 different test sets.

The classification performance of malignant melanoma improves until the contrast factor get the value of 1. Beyond this value, CNN models continued to perform well. On the other hand, classification of benign lesions decreased when contrast factor is larger than 1. This revealed that classification performance of melanoma lesions were not adversely affected by contrast. In addition, classification performance of melanoma images are more robust than benign lesion images at different levels of contrast. When we looked at overall success for both classes, DenseNet worked better than other CNN models in changes of contrast factor. DenseNet gives the best accuracy as 86.37% on this dataset when the contrast factor is 1. ResNet was the second best model with accuracy of 83.7%.

In experiments with different noise levels, it was observed that classification performance of melanoma images were more sensitive to noise. Despite this, benign lesion recognition had a much better performance compared to melanoma. Out of the 4 models, DenseNet121 performed the best on noisy images when both classes are considered. Unlike contrast changes, melanoma classification performance is more sensitive to noise compared to benign lesions. As the images get more noisy, accuracy of melanoma diagnosis decreases considerably. We calculated the model accuracy is 89.03% with DenseNet when standard deviation (σ) was 1.

With the enlargement of the filter kernel used in blurring, the images deteriorate considerably. A rapid decrease is observed with AlexNet and VGGNet model. Between ResNet and DenseNet models; DenseNet has a better success with 88.77% accuracy when standard deviation of filter kernel value was 2 and the Gaussian filter size was 9×9 . When we used VGG16 and AlexNet models for the classification of skin lesions in blur datasets, we get rapidly decreasing accuracy in melanoma lesions however increasing accuracy for the benign lesions. Nevertheless, we observed an oscillating accuracy in the benign lesions with DenseNet121 and ResNet50 models. In this case, false negative results were increasing while the false positives are decreasing.

We believe the models tend towards making benign biased decisions because of the increasing information loss in the lesion edges. The accuracies were at acceptable levels for both classes. The best classification accuracy was observed with DenseNet.

The ruler in the images has decreased the accuracy. We can explain that the number of ruler images were less than the other in both training and test sets. Hence, rulers were not learnt sufficiently for this set. The accuracy of malignant melanoma was better than benign. In this test set, DenseNet and ResNet can be preferred due to their better performance. In ruler test set, DenseNet performed better with average accuracy of 86% compared to other CNN models. In addition, ResNet was closer to DenseNet with accuracy of 85.23%.

In lesions with hair, system produced better results. Due to the high number of lesions with hair in the training set, better training may have achieved for these types of occlusions. DenseNet achieved higher accuracy for this test set. We observed the overall accuracy of the hair dataset as 89.22% with DenseNet121. We expected that the highest accuracy was to be obtained with Test_None. However, better accuracy was obtained on Test_Hair compared to Test_None. Due to the fact that, there were more hairy images compared to mole images. In addition, we can interpret the result that wrinkles or various patterns on the skin should be considered in terms of classification. DenseNet had achieved better performance in the image set with hairs. The lesions without any extra object or distortions are also better recognized with DenseNet which had achieved 88.81% accuracy.

Consequently, considering 6 different test sets, DenseNet had a better classification accuracy regarding image degradation problems.

8.2 Conclusion of Ensemble of Convolutional Neural Networks for Skin Lesion Classification

We endeavored a non-invasive technique to differentiate melanoma and benign melanocytic nevi on the skin. We aimed to design a CAD system that may help clinicians for early diagnosis. Towards this purpose, we used ISIC 2020 dataset. It is an unbalanced dataset. Although there are so many benign lesion images, there are

very few malignant melanoma images. 565 melanoma and 600 benign nevi images were used in this study.

Firstly, we split the dataset as training and testing sets. Then, we divided our training set into 5 different folds and applied data augmentation processes such as flipping and rotation. After that, we trained 11 different CNN models. We classified images in test set by each model separately. Finally, we obtained five class labels (one from each model). We ensembled these CNN models by simple majority voting algorithm.

In this study, not only we proposed a CAD system, but also we analyzed the performances of 11 CNN models for skin lesion classification whose accuracies are listed in Table 8.1.

For analysis, we trained the five folds with same CNN model such as ResNet50 and we had five different trained model. We then applied these models to test set so that we had five different outputs of the models and we fused their outputs by three different approaches.

Table 8.1 : The sorted list of CNN models according to the accuracy.

CNN Model	Accuracy
ResNet18	0.944
ResNet50	0.944
Vgg16bn	0.940
Vgg19	0.940
AlexNet	0.936
SqueezeNet1_1	0.931
SqueezeNet1_0	0.927
DenseNet169	0.927
DenseNet121	0.923
DenseNet161	0.923
DenseNet201	0.919

ResNet models are better than the other models in binary classification of skin lesions. On the other hand, the performance rates of the models differ slightly for the case of classification of lesions separately. VGG16Net and AlexNet had the best accuracy on benign nevi image classification with the ratio of 95%. For malignant melanoma classification DenseNet169, ResNet18 and ResNet50 are better than the other CNN models with equal accuracy of 95.6%.

In medical image processing problems, accurate detection of images that include medical abnormalities is much more important than correct recognition of healthy images. For example, if a malignant lesion is classified as benign, the skin cancer cannot be diagnosed and the necessary treatment cannot be started.

In our study, malignant melanoma was correctly diagnosed with a higher success rate than benign nevi. Accuracy was observed as 98.2% for malignant melanoma and 93.3% for benign nevi recognition. The overall classification accuracy of our CAD system was 95.76%.





REFERENCES

- [1] **International Agency for Research on Cancer, WHO** (2020). Turkey, Globocan 2020, <https://gco.iarc.fr/today/data/factsheets/populations/792-turkey-fact-sheets.pdf>.
- [2] **Foundation, W. C. R.** (2019). Diet, Nutrition, physical activity and Skin cancer, <https://www.wcrf.org/dietandcancer/skin-cancer>.
- [3] **Tschandl, P., Rosendahl, C., and Kittler, H.** (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, *Scientific data*, 5, 180161.
- [4] **International Agency for Research on Cancer, WHO** (2020). Melanoma of Skin, Globocan 2020, <https://gco.iarc.fr/today/data/factsheets/cancers/16-Melanoma-of-skin-fact-sheet.pdf>.
- [5] **Mathworks**, <https://www.mathworks.com/help/deeplearning/ref/alexnet.html>.
- [6] **Feng, P., Lin, Y., Guan, J., Dong, Y., He, G., Xia, Z. and Shi, H.** (2019). Embranchment Cnn Based Local Climate Zone Classification Using Sar And Multispectral Remote Sensing Data, *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, pp.6344–6347.
- [7] <http://karthink.me/journal/summary-densenet.html>.
- [8] **Ramsundar, B. and Zadeh, R. B.** (2018). *TensorFlow for deep learning: from linear regression to reinforcement learning*, " O'Reilly Media, Inc."
- [9] **Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., and De, D.** (2020). Fundamental concepts of convolutional neural network, *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, Springer, pp.519–567.
- [10] **National Cancer Institute** (2009). Skin cancer, <http://cancer.gov/cancertopics/types/skin> Accessed.
- [11] **American Cancer Society** (2021). Melanoma Skin Cancer.
- [12] **Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M.** (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008, *International journal of cancer*, 127(12), 2893–2917.
- [13] **Çalışma Grubu, M.**, (2012), Melanom Yol Haritası, <http://kanser.gov.tr/Dosya/Medya/Haberler/Melanomyolharitasi.pdf>.

- [14] **Testori, A., Ascierto, P., Sileni, V. C., De Lorenzo, F., Pelicci, P., and Rossi, C.** (2014). Report from the Melanoma Independent Board First Melanoma MIB Conference, 21–22 October 2013, *ecancermedicalscience*, 8.
- [15] **Gazioğlu, B. S. A. and Kamaşak, M. E.** (2021). Effects of objects and image quality on melanoma classification using deep neural networks, *Biomedical Signal Processing and Control*, 67, 102530.
- [16] **Jaworek-Korjakowska, J. and Tadeusiewicz, R.** (2013). Hair removal from dermoscopic color images, *Bio-Algorithms and Med-Systems*, 9(2), 53–58.
- [17] **Toossi, M. T. B., Pourreza, H. R., Zare, H., Sigari, M.-H., Layegh, P., and Azimi, A.** (2013). An effective hair removal algorithm for dermoscopy images, *Skin Research and Technology*, 19(3), 230–235.
- [18] **Julie Ann A. Salido, C. R. J.** (2018). Using Deep Learning for Melanoma Detection in Dermoscopy Images., *International Journal of Machine Learning and Computing*, 8(1), 61–68.
- [19] **Norhafiza Hamzah, R. L., Mohd Sapaweh Asli** (2018). Skin cancer Image Detection using Watershed Marker-Controlled and Canny Edge Detection Techniques, *Transactions on Science and Technology*, 5(1), 1–4.
- [20] **Jaworek-Korjakowska, J. and Kleczek, P.** (2018). eSkin: Study on the Smartphone Application for Early Detection of Malignant Melanoma, *Wireless Communications and Mobile Computing*, 2018.
- [21] **Yıldız, O.** (2019). Melanoma detection from dermoscopy images with deep learning methods: A comprehensive study, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 34(4), 2241–2260.
- [22] **Yu, Z., Jiang, X., Zhou, F., Qin, J., Ni, D., Chen, S., ..., and Wang, T.** (2018). Melanoma recognition in dermoscopy images via aggregated deep convolutional features, *IEEE Transactions on Biomedical Engineering*, 66(4), 1006–1016.
- [23] **Zhang, J., Xie, Y., Xia, Y., and Shen, C.** (2019). Attention residual learning for skin lesion classification, *IEEE transactions on medical imaging*, 38(9), 2092–2103.
- [24] **Ozkan, I. A. and Koklu, M.** (2017). Skin lesion classification using machine learning algorithms, *International Journal of Intelligent Systems and Applications in Engineering*, 5(4), 285–289.
- [25] **Seeja, R. and Suresh, A.** (2019). Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM), *Asian Pacific journal of cancer prevention: APJCP*, 20(5), 1555.
- [26] **Li, Y. and Shen, L.** (2018). Skin lesion analysis towards melanoma detection using deep learning network, *Sensors*, 18(2), 556.

- [27] **Albahar, M. A.** (2019). Skin Lesion Classification Using Convolutional Neural Network With Novel Regularizer, *IEEE Access*, 7, 38306–38313.
- [28] **Nie, Y., De Santis, L., Carratù, M., O’Nils, M., Sommella, P., and Lundgren, J.** (2020). Deep Melanoma classification with K-Fold Cross-Validation for Process optimization, *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp.1–6.
- [29] **Jojoa Acosta, M. F., Caballero Tovar, L. Y., Garcia-Zapirain, M. B., and Percybrooks, W. S.** (2021). Melanoma diagnosis using deep learning techniques on dermatoscopic images., *BMC Medical Imaging*, 21(1).
- [30] **Maiti, A. and Chatterjee, B.** (2019). Improving detection of Melanoma and Naevus with deep neural networks, *Multimedia Tools and Applications*, <https://doi.org/10.1007/s11042-019-07814-8>.
- [31] **Mahbod, A., Schaefer, G., Wang, C., Ecker, R., and Ellinge, I.** (2019). Skin lesion classification using hybrid deep neural networks, *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp.1229–1233.
- [32] **Majtner, T., Bajic, B., Yildirim, S., Hardeberg, J. Y., Lindblad, J., and Sladoje, N.** (2018). Ensemble of Convolutional Neural Networks for Dermoscopic Images Classification, *CoRR*, *abs/1808.05071*, <http://arxiv.org/abs/1808.05071>, 1808.05071.
- [33] **Matsunaga, K., Hamada, A., Minagawa, A., and Koga, H.** (2017). Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble, *arXiv preprint arXiv:1703.03108*.
- [34] **González-Díaz, I.** (2017). Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions, *arXiv preprint arXiv:1703.01976*.
- [35] **Menegola, A., Tavares, J., Fornaciali, M., Li, L. T., de Avila, S. E. F., and Valle, E.** (2017). RECOD Titans at ISIC Challenge 2017, *CoRR*, *abs/1703.04819*.
- [36] **Xie, F., Yang, J., Liu, J., Jiang, Z., Zheng, Y., and Wang, Y.** (2020). Skin lesion segmentation using high-resolution convolutional neural network, *Computer Methods and Programs in Biomedicine*, 186, 105241.
- [37] **Goyal, M., Oakley, A., Bansal, P., Dancey, D., and Yap, M. H.** (2019). Skin Lesion Segmentation in Dermoscopic Images with Ensemble Deep Learning Methods, *IEEE Access*.
- [38] **Jafari, M. H., Nasr-Esfahani, E., Karimi, N., Soroushmehr, S. R., Samavi, S., and Najarian, K.** (2017). Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma, *International journal of computer assisted radiology and surgery*, 12(6), 1021–1030.

- [39] **Yuan, Y., Chao, M., and Lo, Y.-C.** (2017). Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance, *IEEE transactions on medical imaging*, 36(9), 1876–1886.
- [40] **Hekler, A., Utikal, J. S., Enk, A. H., Hauschild, A., Weichenthal, M., and Maron, D., R. C. ... Schadendorf** (2019). Superior skin cancer classification by the combination of human and artificial intelligence, *European Journal of Cancer*, 120, 114–121.
- [41] **Marka, A., Carter, J. B., Toto, E., and Hassanpour, S.** (2019). Automated detection of nonmelanoma skin cancer using digital images: a systematic review, *BMC medical imaging*, 19(1), 21.
- [42] **Brinker, T. J., Hekler, A., Utikal, J. S., Grabe, N., Schadendorf, D., and Klode, C., J. ... Von Kalle** (2018). Skin cancer classification using convolutional neural networks: systematic review, *Journal of medical Internet research*, 20(10), e11936.
- [43] **Pathan, S., Prabhu, K. G., and Siddalingaswamy, P.** (2018). Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review, *Biomedical Signal Processing and Control*, 39, 237–262.
- [44] **Harangi, B.** (2018). Skin lesion classification with ensembles of deep convolutional neural networks, *Journal of biomedical informatics*, 86, 25–32.
- [45] **Hosny, K. M., Kassem, M. A., and Foad, M. M.** (2019). Classification of skin lesions using transfer learning and augmentation with Alex-net, *PloS one*, 14(5).
- [46] **Mahbod, A., Schaefer, G., Ellinger, I., Ecker, R., Pitiot, A., and Wang, C.** (2019). Fusing fine-tuned deep features for skin lesion classification, *Computerized Medical Imaging and Graphics*, 71, 19–29.
- [47] **Perez, F., Avila, S., and Valle, E.** (2019). Solo or ensemble? Choosing a CNN architecture for melanoma classification, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.0–0.
- [48] **Rodrigues, D. D. A., Ivo, R. F., Satapathy, S. C., Wang, S., Hemanth, J., and Reboucas Filho, P. P.** (2020). A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system, *Pattern Recognition Letters*, 136, 8–15.
- [49] **Tiwari, S.** (2021). Dermatoscopy Using Multi-Layer Perceptron, Convolution Neural Network, and Capsule Network to Differentiate Malignant Melanoma From Benign Nevus, *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, 16(3), 58–73.
- [50] **Barata, C., Celebi, M. E., and Marques, J. S.** (2017). Development of a clinically oriented system for melanoma diagnosis, *Pattern Recognition*, 69, 270–285.

- [51] **Shorfuzzaman, M.** (2021). An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection, *Multimedia Systems*, 1–15.
- [52] **He, K., Zhang, X., Ren, S., and Sun, J.** (2016). Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.770–778.
- [53] **Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q.** (2017). Densely connected convolutional networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.4700–4708.
- [54] **Krizhevsky, A., Sutskever, I., and Hinton, G. E.** (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp.1097–1105.
- [55] **Simonyan, K. and Zisserman, A.** (2014). Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*.
- [56] **Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K.** (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size, *arXiv preprint arXiv:1602.07360*.
- [57] **Mendonça, T., Ferreira, P., Marques, J., Marçal, A., and Rozeira, J.** (2013). PH2 - A dermoscopic image database for research and benchmarking, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2013*, 5437–5440.
- [58] **Gutman, D., Codella, N. C., Celebi, E., Helba, B., Marchetti, M., Mishra, N. and Halpern, A.** (2016). Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC), *arXiv preprint arXiv:1605.01397*.
- [59] **Rotemberg, V., Kurtansky, N., Betz-Stablein, B., Caffery, L., Chousakos, E., and Codella, D., N. ... Gutman** (2021). A patient-centric dataset of images and metadata for identifying melanomas using clinical context, *Scientific data*, 8(1), 1–8.
- [60] **Ballerini, L., Fisher, R. B., Aldridge, B., and Rees, J.,** (2013). A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions, *Color Medical Image Analysis*, Springer, pp.63–86.
- [61] **DermIS.** Dermatology Information System, [Online]. Available: <https://www.dermis.net>.
- [62] **MED-NODE.** Interactive Dermatology Atlas, [Online]. Available: <http://www.dermatlas.net/atlas/index.cfm>.

- [63] **Zhang, Y. and Wang, C.** (2021). SIIM-ISIC Melanoma Classification With DenseNet, *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, pp.14–17.
- [64] **Aldwgeri, A. and Abubacker, N. F.** (2019). Ensemble of deep convolutional neural network for skin lesion classification in dermoscopy images, *International Visual Informatics Conference*, Springer, pp.214–226.
- [65] **Sarkar, R., Chatterjee, C., and Hazra, A.** (2019). Diagnosis of melanoma from dermoscopic images using a deep depthwise separable residual convolutional network, *IET Image Processing*, 13.
- [66] **Hasan, M., Barman, S. D., Islam, S., and Reza, A. W.** (2019). Skin cancer detection using convolutional neural network, *Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence*, pp.254–258.
- [67] **Almaraz-Damian, J.-A., Ponomaryov, V., Sadovnychiy, S., and Castillejos-Fernandez, H.** (2020). Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures, *Entropy*, 22(4), 484.
- [68] **Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., and Dusza, H., S. W. ... Kittler** (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, pp.168–172.
- [69] **Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., and Gutman, M., D. ... Marchetti** (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic), *arXiv preprint arXiv:1902.03368*.
- [70] **Combalia, M., Codella, N. C., Rotemberg, V., Helba, B., Vilaplana, V., and Reiter, J., O. ... Malvey** (2019). Bcn20000: Dermoscopic lesions in the wild, *arXiv preprint arXiv:1908.02288*.
- [71] **Dodge, S. and Karam, L.** (2016). Understanding how image quality affects deep neural networks, *2016 eighth international conference on quality of multimedia experience (QoMEX)*, IEEE, pp.1–6.
- [72] **Pei, Y., Huang, Y., Zou, Q., Zhang, X., and Wang, S.** (2019). Effects of Image Degradation and Degradation Removal to CNN-based Image Classification, *IEEE transactions on pattern analysis and machine intelligence*.
- [73] **Roy, P., Ghosh, S., Bhattacharya, S., and Pal, U.** (2018). Effects of degradations on deep neural network architectures, *arXiv preprint arXiv:1807.10108*.

- [74] **Zhou, Y., Song, S., and Cheung, N.-M.** (2017). On classification of distorted images with deep convolutional neural networks, *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp.1213–1217.
- [75] **Ghosh, S., Shet, R., Amon, P., Hutter, A., and Kaup, A.** (2018). Robustness of Deep Convolutional Neural Networks for Image Degradations, *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp.2916–2920.
- [76] **Grm, K., Štruc, V., Artiges, A., Caron, M., and Ekenel, H. K.** (2017). Strengths and weaknesses of deep learning models for face recognition against image degradations, *Iet Biometrics*, 7(1), 81–89.
- [77] **Fan, X., Dai, M., Liu, C., Wu, F., Yan, X., and Feng, B., Y. ... Su** (2019). Effect of image noise on the classification of skin lesions using deep convolutional neural networks, *Tsinghua Science and Technology*, 25(3), 425–434.
- [78] **Elmahdy, M. S., Abdeldayem, S. S., and Yassine, I. A.** (2017). Low quality dermal image classification using transfer learning, *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, IEEE, pp.373–376.
- [79] **Kawahara, J., BenTaieb, A., and Hamarneh, G.** (2016). Deep features to classify skin lesions, *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, IEEE, pp.1397–1400.
- [80] **Mahbod, A., Schaefer, G., Wang, C., Ecker, R., and Ellinge, I.** (2019). Skin lesion classification using hybrid deep neural networks, *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp.1229–1233.
- [81] **Codella, N. C., Nguyen, Q.-B., Pankanti, S., Gutman, D. A., Helba, B., Halpern, A. C. and Smith, J. R.** (2017). Deep learning ensembles for melanoma recognition in dermoscopy images, *IBM Journal of Research and Development*, 61(4/5), 5–1.
- [82] **Lopez, A. R., Giro-i Nieto, X., Burdick, J., and Marques, O.** (2017). Skin lesion classification from dermoscopic images using deep learning techniques, *2017 13th IASTED international conference on biomedical engineering (BioMed)*, IEEE, pp.49–54.
- [83] **Harangi, B.** (2017). Skin lesion detection based on an ensemble of deep convolutional neural network, *arXiv preprint arXiv:1705.03360*.
- [84] **Gessert, N., Sentker, T., Madesta, F., Schmitz, R., Kniep, H., Baltruschat, I., ..., and Schlaefer, A.** (2018). Skin lesion diagnosis using ensembles, unscaled multi-crop evaluation and loss weighting, *arXiv preprint arXiv:1808.01694*.

- [85] **Li, K. M.** and **Li, E. C.** (2018). Skin lesion analysis towards melanoma detection via end-to-end deep learning of convolutional neural networks, *arXiv preprint arXiv:1807.08332*.
- [86] **Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., and Chanan, S., G. ... Chintala,** (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library, *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., pp.8024–8035.
- [87] **Sagar, A.** and **Dheeba, J.** (2020). Convolutional neural networks for classifying melanoma images, *bioRxiv*.
- [88] **Mendes, D. B.** and **da Silva, N. C.** (2018). Skin lesions classification using convolutional neural networks in clinical images, *arXiv preprint arXiv:1812.02316*.
- [89] **Rezvantab, A., Safigholi, H., and Karimijeshni, S.** (2018). Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms, *arXiv preprint arXiv:1810.10348*.
- [90] **Dorj, U.-O., Lee, K.-K., Choi, J.-Y., and Lee, M.** (2018). The skin cancer classification using deep convolutional neural network, *Multimedia Tools and Applications*, 77(8), 9909–9924.
- [91] **Harangi, B., Baran, A., and Hajdu, A.** (2018). Classification of skin lesions using an ensemble of deep neural networks, *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, IEEE, pp.2575–2578.
- [92] **Kalouche, S., Ng, A., and Duchi, J.** (2016). Vision-based classification of skin cancer using deep learning, *2015, conducted on Stanfords Machine Learning course (CS 229) taught*.
- [93] **Rashid, H., Tanveer, M. A., and Khan, H. A.** (2019). Skin lesion classification using GAN based data augmentation, *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, pp.916–919.
- [94] **DeVries, T.** and **Ramachandram, D.** (2017). Skin lesion classification using deep multi-scale convolutional neural networks, *arXiv preprint arXiv:1703.01402*.
- [95] **Mahbod, A., Schaefer, G., Wang, C., Dorffner, G., Ecker, R., and Ellinger, I.** (2020). Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification, *Computer methods and programs in biomedicine*, 193, 105475.
- [96] **Chaturvedi, S. S., Tembhurne, J. V., and Diwan, T.** (2020). A multi-class skin Cancer classification using deep convolutional neural networks, *Multimedia Tools and Applications*, 79(39), 28477–28498.

- [97] **Akanbi, O. A., Amiri, I. S., and Fazeldehkordi, E.,** (2015). Chapter 5 - Implementation and Result, **O.A. Akanbi, I.S. Amiri and E. Fazeldehkordi**, editors, *A Machine-Learning Approach to Phishing Detection and Defense*, Syngress, Boston, pp.55–73.
- [98] **Hasan, M. K., Elahi, M. T. E., Alam, M. A., and Jawad, M. T.** (2021). DermoExpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation, *medRxiv*.
- [99] **Wei, L., Ding, K., and Hu, H.** (2020). Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network, *IEEE Access*, 8, 99633–99647.
- [100] **Harangi, B., Baran, A., and Hajdu, A.** (2020). Assisted deep learning framework for multi-class skin lesion classification considering a binary classification support, *Biomedical Signal Processing and Control*, 62, 102041.
- [101] **Kassem, M. A., Hosny, K. M., and Fouad, M. M.** (2020). Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning, *IEEE Access*, 8, 114822–114832.
- [102] **Almaraz-Damian, J.-A., Ponomaryov, V., Sadovnychiy, S., and Castillejos-Fernandez, H.** (2020). Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures, *Entropy*, 22(4), 484.
- [103] **Tan, T. Y., Zhang, L., and Lim, C. P.** (2020). Adaptive melanoma diagnosis using evolving clustering, ensemble and deep neural networks, *Knowledge-Based Systems*, 187, 104807.
- [104] **Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., and Ozair, Y., S. ... Bengio** (2014). Generative adversarial nets, *Advances in neural information processing systems*, 27.
- [105] **Adadi, A. and Berrada, M.** (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI), *IEEE access*, 6, 52138–52160.
- [106] **Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B.** (2019). Definitions, methods, and applications in interpretable machine learning, *Proceedings of the National Academy of Sciences*, 116(44), 22071–22080.
- [107] **Van der Velden, B. H., Kuijf, H. J., Gilhuijs, K. G., and Viergever, M. A.** (2021). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis, *arXiv preprint arXiv:2107.10912*.
- [108] **Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A.** (2016). Learning deep features for discriminative localization, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.2921–2929.

- [109] **Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., and Qian, W.** (2019). An interpretable ensemble deep learning model for diabetic retinopathy disease classification, *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, pp.2045–2048.



CURRICULUM VITAE

Name Surname: Bilge Süheyla AKKOCA GAZİOĞLU

EDUCATION:

- **B.Sc.:** 2010, Karadeniz Technical University, Faculty of Architecture and Engineering, Computer Engineering Department
- **M.Sc.:** 2014, Istanbul Technical University, Graduate School, Computer Engineering Department

PROFESSIONAL EXPERIENCE AND REWARDS:

- Worked as a computer engineer in BAMA Technology between 2010-2011
- Worked as a research assistant at Istanbul Technical University Computer Engineering Department between 2011-2021

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **GAZİOĞLU, B.S.A., KAMAŞAK, M.E.,** "Effects of objects and image quality on melanoma classification using deep neural networks", Biomedical Signal Processing and Control, 2021.