ANALYSIS REPORT

Similarity document :

 3%

Similarities section 2 :

 4%

ANALYZED ON THE ACCOUNT

Last name :	MEDDAH
First name :	Ishak Hibat Allah Abdelwaheb
Email :	i.meddah@esi-sba.dz
Folder :	Default folder

DOCUMENT INFORMATION

Author(s) :	not available
Title :	Head-master.pdf
Description :	not available
Analysed on :	06/21/2022 6:44 PM
Document ID :	u9k7po6e
File name :	head-master.pdf
File type :	pdf
Word count :	5 663
Character count :	39 004
Original file size (kB) :	6 277.25
Upload type :	manual submission
uploaded on the :	06/21/2022 6:30 PM

FOUND SOURCES

Highly probable sources :	7 sources
Less probable sources :	0 source
Accidental sources :	25 sources
Ignored sources :	0 source

SIMILARITIES FOUND IN THIS DOCUMENT/SECTION

Matching similarities :	3%
Assumed similarities :	<1%
Accidental similarities :	<1%

TOP PROBABLE SOURCES - AMONG 7 PROBABLE SOURCES

Sources	Similarity
1. Document: 9vytnx4a - belongs to another user	1%
2. Source Compilatio.net gibxh4ro	<1%

HIGHLY PROBABLE SOURCES

7 Sources	Similarity

1.	Source Compilatio.net 9vytnx4a	1%
2.	**www.sciencedirect.com**/.../pii/S2666351121000498	<1%
3.	Source Compilatio.net gibxh4ro	<1%
4.	**www.healthyskinworld.com**/.../is-basal-cell-skin-cancer-dangerous	<1%
5.	**www.sciencedirect.com**/.../pii/S2352914819302047	<1%
6.	**www.skincancer.org**/.../skin-cancer-information/melanoma	<1%
7.	**www.sciencedirect.com**/.../pii/S2214785320354766	<1%

LESS PROBABLE SOURCES

0 Source

ACCIDENTAL SOURCES

25 Sources

		Similarity
1.	Source Compilatio.net it5792om	<1%
2.	Source Compilatio.net 5ad78irj	<1%
3.	Source Compilatio.net devy48	<1%
4.	Source Compilatio.net psxm21r4	<1%
5.	Source Compilatio.net k2udgfn4	<1%
6.	Source Compilatio.net sfydqcx7	<1%
7.	Source Compilatio.net fyqpbcwv	<1%
8.	**towardsdatascience.com**/.../model-agnostic-met...model-4f10787ef504	<1%
9.	Source Compilatio.net k7brzhp6	<1%
10.	Source Compilatio.net yor8d5n7	<1%
11.	Source Compilatio.net cemxf53p	<1%
12.	Source Compilatio.net i8r6hdlz	<1%
13.	Source Compilatio.net 7an5e4sd	<1%
14.	Source Compilatio.net os4m68nc	<1%
15.	Source Compilatio.net v8nguyd1	<1%
16.	Source Compilatio.net jzdcw76v	<1%
17.	**towardsdatascience.com**/.../what-are-model-agn...thods-387b0e8441ef	<1%
18.	Source Compilatio.net 1fqm5ukc	<1%
19.	Source Compilatio.net g4db9x7v	<1%
20.	**www.sciencedirect.com**/.../pii/S2213909520300033	<1%
21.	**www.sciencedirect.com**/.../pii/S0306987721002231	<1%
22.	**www.sciencedirect.com**/.../pii/S0306987721002231	<1%
23.	Source Compilatio.net 34ixzmvn	<1%
24.	Source Compilatio.net t6phe8m9	<1%
25.	**pubmed.ncbi.nlm.nih.gov**/.../34688214	<1%

IGNORED SOURCES

TEXT EXTRACTED FROM THE DOCUMENT

Caption : Texts within quotation marks

third function [33]) this architecture of neural networks is specialised in processing image data and it is built on three primaty layers Convolution layer, pooling layer and fully connected layer [9] 24 Convolution layer this layer does most of the hard work by identifying and extracting the features, this is done by applying a filter of random size to blocks of the input image using the dot product between matrices pooling layer after the feature extraction resulting from the Convolution layer we need to Simplify (by reducing a bloc of values to a single value) the image for easy learning, there are 2 pooling operations max pooling and average pooling fully connected layer it operates on a flatened input, where each input is connected to all of the neurons, it is usualy found at the and of the network connecting the hidden layers to the output which help in optimizing the class scores Figure 2.3: Deep Learning in Computer Vision [9] 2.5.4 applications of computer vision there are alot of indsutries using computer vision and these are just a few examples [9] medical imaging it helps medical professionals interpret faster and diagnose abnormalities. law enforcement and security like in surveillance and authentication self driving machines like cars and robots gaming augmented reality and virtual reality pattern recognition 2.5.5 some technologies of computer vision because of the wide utility of computer vision and its benefits there are alot of libraries and frameworks that facilitates alot of the hard and repeated tasks, here we mention a few of them [9] openCV a python library for computer vision, 25 • super easy to use, • a huge library of image processing algorithms, • open source, • works with GPUs Tensorflow made by Google and one of the most popular machine learning frameworks • with a wide rane of machine/deep learning algorithms, • open source, • GPU configured PyTorch made by facebook a neural network framework, • used alot by researchers, • open source, • works with GPUs Caffe a deep learning framework developed by Berkeley AI Research • open source • c++ based • easy to use • fats execution 26 Chapter 3 State Of The Art 3.1 Introduction In this chapter we are going to present some of the recent work and research that was done regarding skin cancer detection and classification using machine learning , we are going to explore the various methods, tools, new ideas and challenges that was handeled by researchers for the hope of getting a clear understanding

of the problem and how to

go about solving it depending on each one’s conditions, requirements and goals.

3.2

skin cancer detection and classification using machine learning

proposed methodology the proposed methodology in this article [35] uses a 6 step

process (input data - preprocessing - segmentation - feature extraction - classification - output data)

input data dermoscopic images from the ISIC (International Skin Imaging Collaboration) 2019 challenge containing 8 classes of skin lesions, and for simplisity reasons

only 800 images out of 25000 is used.

preprocessing because of the heteroginity of the input data a preprocessing step is

required to inhance the quality of images and remove irrelevant parts. the main

technices used here are gray scale conversion and the application of the Gaussian

and median filter for noise removal and enhancement, and for the unwanted hair

they applied the Dull Razor method (a preprocessing algorithm), as shown in figure

3.1

segmentation segmnetation is used to extract the region of interest and for that they

used a k-means clustering algorithm as shown in figure 3.2

feature extraction for this they used 2 well know methods, ABCD method and GLCM.

ABCD is used in dermatological applications and diagnosis for skin lesions such as

melanomas and it is the abreviation of Asymmetry, Border, Color and Diameter.

Grey Level Co-occurrence Matrix (GLCM) is used for texture analysis, other features are also used in addition to these 2 methods for further classification such as

Autocorrelation, correlation, Standard vector...etc

27

classification for classification they used MSVM (Multi-class Support vector machine)

machine learning algorithm, they used trainning and testing ratios of 70:30 and

obtained an acuuracy of 96.25% and the confusion matrix shown in figure 3.3

Figure 3.1: Preprocessing: (a)Dull razor image, (b) Gray scale image, (c) Gaussian filter,

(d) Median filter.

Figure 3.2: Segmentation: (a) Image labelled by cluster index, (b) Objects in cluster 1,

(c) Objects in cluster 2, (d) Objects in cluster 3.

3.3

Finding reduced Raman spectroscopy fingerprint

of skin samples for melanoma diagnosis through

machine learning

This article [36] uses a new non invasive approach to classify malignant and benign

tumors, and that is by using Raman spectral data instead of images, Raman Spectroscopy

is a way to analyse the chemical structure using light and vibrational energy modes of

molecules [37]

data and method

dataset: for the dataset they brought 33 benign and 51 malignant smaples and

cut them into regular cuts of 2mm3 , a layser was used to excite the samples to

collect the Raman signals using special tools after this they aquired 436 Raman

28

Figure 3.3: Confusion Matrix

29

spectra (spectrum graphs y=f(x) where x is frequency or wavenumber cm 1 and y

is the intensity of scattered light). and they focused on the biological fingerprint

spectral region from 800 to 1800 cm 1

Fluorescence background data pre-process: Fluorescence is a radiation that

is emitted by molecules after interacting with electromagnetic radiation and this

could overshdow and disturb the study of Raman spectra, to deal with this noise

they used a low frequency laser to lower the probability of fluorescence emissions

and by this they could jump the preprocessing step.

feature extraction they divided the obtained spectrums into subsequences (local spectrums) and extracted some statistical measures from it such as arithmetic mean,

standard deviation, derivative ...etc

results and discussion these statistical features were then given to a machine learning

classification algorithm, a complex dicision tree impimented using lightGBM (open

source software), other algorithms were also used such as K-nearest neighbors and

XGBOOST (Extreme Gradient Boosting an open source software) but the best

performance was obtained using lightGBM. further research led them to only use

the derivative as a feature and a spectral region from 896 to 1039 cm 1 because these

two were proved to have the most discriminative information between malignant and

benign tumors and by this they obtained a high performant model (AU C 0.97)

shown in figure 3.4

Figure 3.4: ROC

3.4

Skin cancer**detection: Applying a deep learning**

based model driven architecture in the cloud for

classifying dermal cell images

Summary

in this paper [38] the researchers are presenting a model driven approach to develop

30

deep learning algorithms for detecting skin cancer by using a tool called DLS (deep

learning studio) which is a software that allows you to build deep learning algorithms

without being a specialist in programming languages, it presents a simple drag and

drop interface for building models it also commes with desktop / cloud versions

and community / enterprise editions with multi-GPU trainning and the possibility

to obtain the code of the model, download the model and host it as a REST API

(Representational state transfer Application programming interface), the interface

dashboard is shown in figure 3.5

Advantage

the advantage of this non programatic approach is for researchers and practitionners to be able to create and test there own models without the need for prior

programming knowledge

Application and Results

and then they procede using this tool DLS to show its efficacy and ease of use, they

have built and tested 5 models using famous architectures squeeznet, densenet, and

inception v3 with model1 aquiring an AUC of 99.77%

Figure 3.5: DLS Interface

3.5

The impact of patient clinical information on automated skin cancer detection

In this work [39] the researchers propose a new idea, which is the use of clinical information in addition to the image dataset and the study of this addition’s effect on the deep

learning model’s performance

dataset

to build their hybrid dataset, they proposed a mobile application given to doctors

31

and students to help collect the necessary data from Dermatological Assistance Program (PAD) dataset at the Federal University of Esp´rito Santo (UFES), which consists of images of the lesion, their clinical diagnosis and 8 clinical information based on common questions that dermotologists ask:

• age

• part of the body where the lesion is located,

• if the lesion itches,

• bleeds or has bled,

• hurts,

• has recently increased,

• has changed its pattern,

• and if it has an elevation

a total of 1612 images of 6 lesions

because the image dataset is imbalanced they used multiple strategies to overcome that such as, transfer learning (refining a pretrained model on there dataset) , data augmentation, horizontal and vertical rotations, adjusting brightness...etc, and for the clinical data they used one-hot encoding (converting categorical data to augment the performance) which transformed the 8 features collected to an array of 28 values

trainning

they used 4 CNN architectures VGGNet-13/19-bn, ResNet-50/101, MobileNet, GoogleNet now a problem arised when trying to combine (by concatenation) clinical data with image features extracted by the CNN feature extractor because image features are far more great in size then clinical data, this imbalance is not good for the trainning and classification because the effect of image features will be greater then the clinical data, that is why they they implimented an NN feature reducer on the extracted image features before combining it with the clinical data as shown in figure 3.6 [clinical-image.png] and the classifier is another neural network that assigns the probabilities for each skin lesion

testing the effect of adding clinical data

they executed 2 scenarios for that, 1 using models trained only with images, 2 using models trained with images + clinical data then they calculated multiple performance metrics accuracy , balanced accuracy , weighted precision , weighted recall , weighted F1 score and area under the curve and they found almost all models was improved by 7% in almost all metrics and the best model ResNet-50 presented an AU C 95.8%

conclusion

clinical information does make a difference when trainning ML models to classify skin cancer

32

Figure 3.6: Model

3.6

An artificial neural network based detection and

classification of melanoma skin cancer using hybrid texture features

in this work [40] they try to combine multiple texture features from famous methods

such as ABCD, GLCM and LBP (local binary pattern) and pass all of these features to

an ANN (artificial neural network)for learning

dataset

they prefered to use images captured using a dermatoscope because of their quality

over images captured using a phone or normal camera and they have obtained

these images by combinnig 2 datasets: ISIC archive dataset (jpg format) and PH2

dataset (a dermoscopic image dataset in BMP format) they formed a unified dataset

containning 1940 benign and 1448 malignant lesion images

preprocessing

because the images are obtained from various sources, they needed to process them

to standardized them in size, shape, format ...etc and also to remove noise and enhance image quality using enhancement algorithms such as histogram equalization

process that increases image contrast, and to remove body hair from the images

using Maximum Gradient Intensity (MGI) algorithm

image segmentation

for better analysis and to remove unwanted parts they segmented the images to

keep only the lesion area and for that they used a segmentation method called

Otsu’s Thresholding

feature extraction

they used ABCD (Asymmetry, Border, Color, Diameter), GLCM (energy, contrast,

correlation, homogeneity) and LBP (local binary pattern used for textural analysis)

as features to train there neural network

classification

a feed-forward neural network with backprobagation mechanism is used with the

input layer receiving the extracted features and a hidden layer of 100 neurons and

an output layer for the final result (1 is malignant and 0 is benign) with baises

33

and weights initialised randomly, Levenberg-Marquardt trainning and optimization

functions are used and while the performance function being Mean Square Error

and 2 activation functions ”tansig” for the hidden layer and ”purelin” for the final

output the structure of the ANN is shown in figure 3.7 [ann.png]

evaluation

for the evaluation of there classifier they calculated accuracy, specificity, sensitivity

and precision shown in figure 3.8 [evaluation.png] where all the mesures are > 97%,
and further more they also studied the effect of each feature on the discrimination
process between benign and malignant lesion and they found that the minimum
sensitivity per single feature is 69%, minimum specificity per single feature is 73
and minimum accuracy per single feature is 71% which goes to show that all the used
features are playing an important role in the classification process and lastly they
did a comparative evaluation between their work and previous works on the basis
of extracted featues which showed that more featues implies higher performance
rates, an example of that is the accuracy of previous
works using a combination of

some but not all features in (ABCD, GLCM, LBP) always presented an accuracy
< 97%
in conclusion: the use of hybrid features provided a higher performant model in the
detection and classification of benign and malignant melanoma skin cancer
Figure 3.7: ANN Structure
34
Figure 3.8: Evaluation Mesures
3.7
Interpretable deep learning systems for multiclass segmentation and classification of non-melanoma
skin cancer
this article [41] talks about acheiving interpretability in deep learning based systems, and
the reason for this is that traditional machine learning models do outpreform professionals
in some scenarios but we cant explain their output because they are like black boxes we
dont know what is actualy going on inside or why the model chose this output instead of
another output, so it can not be trusted in high stake situations, for that interpretability
of models became a thing in recent research articles, and there are mainly 2 ways in which
we can attain interpretability, I.) we can use Model-Agnostic Methods for Interpreting
any Machine Learning Model:

• like permutation feature importance,

• Partial Dependence Plots (PDPs),
• Individual Conditional Expectation (ICE) plots,
• global surrogate models,
• Local Interpretable Model-agnostic Explanations (LIME)
• Shapley Additive Explanations (SHAP)

to try and explain our model [42] which are statistical and visual ways used to understand
a model, II.) there is another way which is the one used in this article and that is
“naturally interpretable models”, which can be defined as models that try to solve the

problem the way a human would, which means in the case of skin cancer, analysing the whole tissue (the same way a doctor would) and not just cancerous regions of interest and we can acheive this with semantic segmentation methods

dataset

MyLab Pathology provided them with there pre-existing images on non-melanoma

35

cancers, which was taken by a microscope (one image of a cancer is the result of multiple microscopic images concatenated together) for punch, shave and excision biopsies (shave: the surface of the skin is removed with a sharp knife, punch: a round small part of the skin is removed) which meant high resolution images (1px=0.67µm) and each image was annotated by a pathologist to indicate important tissue section in the discrimination process, any imbalance of classes was solved using augmentation (rotation and flipping...etc)

models

whole image segmentation

input: microscopic image see figure 3.9a

ouput: h x w x 12 (12 probability distribution maps), see figure 3.9b

the different tissue sections were colored

1.

2.

3.

4.

5.

6.

7.

8.

9.

10.

11.

12.

Glands (GLD)

Inflammation (INF)

Hair Follicles (FOL)

Hypodermis (HYP)

Reticular Dermis (RET)

Pap-illary Dermis (PAP)

Epidermis (EPI)

Keratin (KER)

Background (BKG)

BCC (Basal cell carcinoma)

SCC (Squamous cell carcinoma)

IEC (a very early treatable form of skin cancer)

to be fed to the segmentation model to train on semantic segmentation, this

model was created using a combination of U-net-like architecture (U-net: a famous CNN architecture for biomedical image segmentation) and a pretrained

headless ResNet50 network. now because of the high resolution of the microscopic images they were fed to the model in parts of 256x256 and 512x512

pixels

whole image classification

input: output of segmantation (h x w x [12 images]) was given to classification

output: 4 classes Healthy, BCC, SCC and IEC

the output of whole image segmantation (which was a probability distribution

for each pixel on the 12 tissue classes) was fed to a CNN to train as a classifier

using Adam optimizer (used to accelerate the gradient descent algorithm) and

a learning rate of 0.0001, with a ratio of 80:10:10 for trainning, validation and

testing

results and discussion

the segmentation model achieved a per-pixel accuracy of 86% and overall class

accuracy of 85%, they found that downgrading the images size before trainning to

36

10 times less increases the accuracy but only by a little bit 2% which isnt much

but this information is still usefull because it means that we can use low resolution

images and still get a high performant model with less computational power

the classification model achieved an accuracy of 93.6% over the 4 classes compared

to other algorithms trained with the same data such as (Random Forest 87.2%,

KNN 80.9%, Single-layer Perceptron 85.1%)

conclusion

they showed that in order to build an interpretable model for skin cancer detection and classification you need to train your model the same way a real doctor

would try to diagnose the skin cancer , and they did that by feeding and trainning

the algorithm with the same data a dermotologist would use for diagnosis without

ignoring any thing such as haire, sweat glands ...etc, in the end they obtained a

high performant model that is interpretable (which means that when a doctor sees

the classification of the algorithm he can understand why it chose that classification) which will increase the possibility to use this approach in real life high stake

scenarios, further more, because they used a diverse dataset said there algorithm

can be used for more routine work that a dermatologist would do such as assessing

aggressiveness, depth, direction of growth and even calculating surgical margins (to

know how much tissue to remove to guarantee that all cancerous cells are removed)

(a) Input

(b) Output

Figure 3.9: Whole segmentation model input and output

3.8

Comprehensive Comarative Information

In this section we are going to present a comprehensive comparative information for the various methods tested on benchmark datasets and there evaluation metrics that were extracted from previous letirature work, we present these tables as a recap since all the methods pass through the same steps (dataset, preprocessing, feature extraction and prediction), we present a table 3.1 that recaps the methods mentioned above and 2 more tables with other methods 3.2 3.3, and it is also worth mentionning that the best methods used in the last 6 years (2016, 2021) using KPI-accuracy are SVM in machine

37

learning and CNN in deep learning and the utilization shares are 54% for ML methods and 46% for DL methods [43]

Beginning

Method and Dataset

ResNet/

Hallym

dataset

19,398

instances are used

Whale algorithm applied to optimize the

CNN model/ Around

22,000 images of Dermquest and DermIS

dataset were used

ResNet/ Clinical images/ Clinical data are

included

ResNet/ 1279 dermoscopic instances are

used

ResNet/

HAM10000

dataset, ISIC archive11,444 dermatoscopic

instances are used

ResNet/

Dermnet

dataset provides an instance of 23 categories

of skin diseases

Dragonfly

optimized

DNN model is assessed using existing

techniques

such

as

Support Vector Machine, ANN, and to

display the efficiency

of the system diverse

evaluation

criteria’s

are accuracy,sensitivity,

and

specificity

are

considered

Results

For Basal Cell Carcinoma (BCC)/ 96% for

asian dataset/ 90% for

the instances of caucasianFor Melanoma –/

96% for the instance

of asian/ 88% for caucasian dataset

Achieved :/ Specificity

98%/ Accuracy- 94%/

Sensitivity-97%/ PPV90%

Limitations

Model performance depends on patient ethnicity

Achieved :/ 67.1% for

clinical images/ 78.8%

for clinical data and images

Achieved 89.2% for

ResNet without dermatologist/ 94.3% for

ResNet- with dermatologist

Mean accuracy of/

Physician 42.94/ CNN

model – 81.59/ Fusion

model – 82.59

Achieved 97.1%

Missing values are not handled in

clinical data/ Biopsy images are not

included

Achieved/ Sensitivity

– 84%/ Specificity99.5%/ Accuracy –

98.5%

38

Results are unsatisfactory

non–Melanoma cases

for

Clinical information to be included/

No standard evaluation criteria exist to measure classification efficacy

Model outperformed for the trained

dataset, but performance degraded

for other datasets.

Able to identify few skin disease

like/ Acne , Rosacea, Hemangioma,/ Psoriasis, Seborrheic Dermatitis

Result generated for Melanoma is

inadequate

Method and Dataset

CNN model used to

Categorized the skin

cancer into Melanoma,

nevi, basal cell carcinoma/ 11,444 dermoscopic

images

HAM10000

dataset,

ISIC archive was used

GoogleNet

CNN

model/

4800

clinical images

DenseNet,

ResNet

model results compared

with

Dermatologist/

10,135

dermoscopy

images of HAM10000:

10015, PH2: 120 data

sets are used

Continuation

Results

89.2% of specificity and

56.5% of sensitivity attained by the dermatologists/ Whereas the

CNN model achieved

98.8%.

of sensitivity,and specificity

Achieved/ Sensitivity

96.3% Specificity 59.5%

For Melanoma and

BCC – achieved94.40%

for ResNet, 99.30%

for DenseNet whereas

Dermatologist achieved

– 82.26% and 88.82%

accuracy

Limitations

Model biased and failed to classify

skin lesion properly/ For melanoma

images CNN model diagnosed it

as nevi whereas dermatologists diagnosed it as melanoma/ model

achieved the lowest specificity 94.2%

for melanoma class

Model behaviour depends on skin

colour tone, and Model biased, sensitivity and specificity dropped for

Caucasian patients/ No standard

evaluation criteria exist to measure

classification efficacy

Model biased for the diverse

dataset/ Model behaviour depends

on skin tone colour

Table 3.3: Another comparative table of latest methods used in skin lesion detection and

there limitations [12]

39

Method

classification using MSVM

Dataset

800 dermoscopic

images ISIC of 8

lesion

classification

436

Rausing

light- man

spectra

GBM (complex (x:frequency,

dicision tree)

y:intensity

of

scattered light)

for benign and

malignant

CNN

famous Kaggle

image

architectures

dataset

(impimented

using

deep

learning studio)

classification

images

(1612

using multiple image of 6 leCNN architec- sions)+ clinical

tures

info

binary classifica- hybrid features

tion ANN

ABCD, GLCM,

LBP

from

dermoscopic

ISIC+PH2

datasets (1940

benign + 1448

malignant)

segmentation to MyLab Pathol12 tissue classes ogy

provided

CNN

(U-net, access to their

Resnet)

and pre-existing

CNN classifica- collection of skin

tion (healty and cancer slides

3 lesions)

Results

accuracy 96.25%

AUC > 97%

AUC 99%

AUC 95.8% and clinical added

a 7% to almost all performance

metrics

performance mesures all > 97%,

effect of each feature > 69%

93.6% accuracy compared to

(Random Forest 87.2%, KNN

80.9%, Single-layer Perceptron

85.1%)

Table 3.1: A comparative table of the methods mentioned above in this article

40

Method

DCNN

Dataset

PH2/ISBI

2016/ISBI 2017

GLCM features ISIC

to an SVM

hybrid adaboost Skin

Cancer

SVM

and

Benign

Tumor

Image

Atlas-Contents

ABCD featues PH2

to an SVM

FCRN architec- ISIC

ture

ANN

CNN

Results

98.4% on PH2 dataset, 95.1% on

ISBI dataset and 94.8% on ISBI

2017 dataset

95% (Accuracy) 90% (sensitivity)

85% (specificity)

91.7% (Accuracy) 94.1%(sensitivity)

88.7%(specificity)

0.83%(Kappa)

90.63% (Accuracy) 95% (sensitivity) 83.33%(specificity)

0.912%(AUC) 0.857% (Accuracy)

0.490%(sensitivity) 0.961%(specificity) 0.729%(average precision)

ISIC

74.76% (Accuracy) 57.56% (validation loss)

Large collection 75.2 (Accuracy) 0.71 (validation

of Multi-Source loss)

Dermatoscopic

Images

Table 3.2: A comparative table of latest methods used in skin lesion detection.

DCNN(Deep convolutional neural network), FCRN (Fully Convolutional Residual Networks) [13]

41

Chapter 4

Results and Discussion

42

Chapter 5

Our Contribution

5.1

Introduction

Melanoma is a type of skin cancer, develops in the cells (melanocytes) that produce melanin

— the pigment that gives your skin its color, The exact cause of all melanomas isn’t clear, but exposure to ultraviolet (UV) radiation from sunlight increases your risk of developing melanoma. [44]

melanoma is more dangerous because of its ability to spread to other organs more rapidly if it is not treated at an early stage. [45]

At present, CNN has achieved very good performance in the field of computer vision, such as object detection, image recognition, classification, etc.

Convolutional Neural Network (CNN)

Main source Document: gibxh4ro - belongs to another user
User: cku127 - confidential Group: ejnw6 - confidential <1%

is a type of deep learning model for processing data that has a grid pattern, such as images, which is designed to automatically and adaptively learn

Main source Document: 9vytnx4a - belongs to another user
User: fikxy - confidential Group: fin23 - confidential 1%

spatial hierarchies of features. CNN is a mathematical construct that is typically composed of three types of layers (or building blocks): convolution, pooling, and fully connected layers. The first two, convolution and pooling layers, perform feature extraction, whereas the third, a fully connected layer, maps the extracted features into final output, such as classification.

A convolution layer plays a key role in CNN, which is composed of a stack of mathematical operations, such as convolution, a specialized type of linear operation. [46]

Because of the difficulty of detecting melanoma cancer in an ordinary way CNN is used to classify melanoma skin cancer.

Research on the classification and detection of melanoma cancer by various methods has been carried out. In 2016 there was a paper entitled ”Deep Residual Learning for Image Recognition” using the ResNet architecture. The paper was a winner at the 2015 ILSVRC (Imagenet competition). [47]

5.2

Proposed Convolutional Neural Network Model

The main aim of this implementation is to detect melanoma skin cancer through RGB images, to achieve this, we build a deep learning model that is capable of extracting features from the given dataset.

After delving into many articles and studies, we have found that the best convolutional neura network model we can suggest in this case is resnet50 and so we are going to implement i from scratch. as shown in figure 5.1

43

Figure 5.1: proposed architecture which we have used for melanoma recognition

5.3

Dataset (MNIST- HAM10000)

The ISIC archive is the largest public database for dermatoscopic image analysis research, and

where the original HAM10000 was made available. [10]

The HAM10000 dataset is composed of 10.015 dermatoscopic images of pigmented skin

lesions. The data was collected from Australian and Austrian patients. Two institutions participated in providing the images: Cliff Rosendahl in Queensland, Australia, and Medical University of Vienna, Austria. According to the authors, seven classes are defined on this dataset

where some diagnoses were unified into one class for simplicity. Information regarding patient

age, sex, lesion location and diagnosis is also provided with each image. [10]

The dataset has been collated and published by Tschandl, P., Rosendahl, C. & Kittler,

H. [10] A sample of each type of skin lesion present in the dataset is demonstrated in the figure

5.2. and the distribution of lesions is show in figure 5.3

5.4

Pre-processing

Before starting the model training process we need to process the dataset, as we learned earlier

the dataset consists of around 10015 labeled images for 7 different types of skin lesions, but in

our case, we want to get images classified on only two types of skin lesions (Melanoma and Not

melanoma). We do this in several steps:

Data cleansing:

In this step, we remove unused and damaged data, also repair data that is incorrectly

formatted.

Data separation:

After cleansing the data set, we separate the data set into two types of skin lesions by

changing the data label for the non-melanoma types to non-melanoma and we keep the

data label for the type of melanoma as it is.

Data balancing:

When reclassifying the data set, we notice that the data set is numerically unbalanced. To

solve this problem, we increase the number of images of the melanoma type by rotating,

44

Figure 5.2: A sample of each type of skin lesion [10]

45

Figure 5.3: This count plot helps to understand the distribution of the data. [10]

cropping and scaling. As for the non-melanoma type, we reduce the number of images by

randomly selecting a specified number of images.

Image resizing:

In this step, we reduce the image size to 75*100 to speed up the training process of the

deep learning model. data splitting : Before the data set becomes usable, we divide it

into two parts, the first part is the training set with 80 percent, and the second part is the test set with 20 percent

The diagram 5.4 helps to understand these steps

Figure 5.4: Pre-processing

5.5 Experimental results

To judge the performance of the model for the task of predicting skin lesions, we use several evaluation metrics to evaluate our model. This is because the model may perform well using one

measurement from one evaluation metric, but may perform poorly using another measurement from another evaluation metric. Using evaluation metrics are critical in ensuring that our model is operating correctly and optimally.

When the model was trained for 30 epochs, it was observed that the accuracy for both the training and test data started with rather large values and continued to increase small from epoch 4 until it reached its peak in epoch 30, where the test accuracy reached 93 percent and the training accuracy was 97 percent.

The plot for the accuracy and loss obtained during the training and testing process is shown in Figure 5.5

Figure 5.5: Accuracy and Loss

The table 5.1 also includes several other measurements that we used in evaluating our model

Classes	Precision	Recall	F1-score	Support
Non-melanoma	0.95	0.93	0.94	1293
Melanoma	0.93	0.95	0.94	1226

Table 5.1: Evaluation Mesures

Bibliography

[1] Skin, cleveland 10978-skin.

clinic.

https://my.clevelandclinic.org/health/articles/

[2] Hair and human skin anatomy.

3190201-hair-and-human-skin-anatomy.

https://www.vecteezy.com/vector-art/

[3] Skin barrier basics for people with eczema, national eczema association.

nationaleczema.org/what-is-my-skin-barrier/.

https://

[4] Vitamin d and the skin: Focus on a complex relationship. https://www.ncbi.nlm.nih.

gov/pmc/articles/PMC4642156/.

[5] What is a mutation? https://www.yourgenome.org/facts/what-is-a-mutation.

[6] Chromothripsis. https://en.wikipedia.org/wiki/Chromothripsis.

[7] Differences between a malignant and benign tumor, verywell health. https://www.

verywellhealth.com/what-does-malignant-and-benign-mean-514240.

[8] Yulia Gavrilova. Artificial intelligence vs. machine learning vs. deep learning: Essentials,

serokell.io. https://serokell.io/blog/ai-ml-dl-difference, 2020.

[9] Vihar Kurama. Ml-based image processing, nanonets. https://nanonets.com/blog/

machine-learning-image-processing/, 2021.

[10] JULIANA NEGRINI DE ARAUJO. Ham10000: Analysis and model comparison. https:

//www.kaggle.com/code/jnegrini/ham10000-analysis-and-model-comparison/

notebook, 2021.

[11] Cancer, world health organization.

detail/cancer.

https://www.who.int/news-room/fact-sheets/

[12] S. Naresh Kumar and B. Mohammed Ismail. Systematic investigation on multi-class skin

cancer categorization using machine learning approach. Materials Today: Proceedings, 12

2020.

[13] Tanzila Saba. Recent advancement in cancer detection using machine learning: Systematic

survey of decades, comparisons and challenges, 9 2020.

[14] Elin Jørgensen, Giulia Lazzarini, Andrea Pirone, Stine Jacobsen, and Vincenzo Miragliotta.

Normal microscopic anatomy of equine body and limb skin: A morphological and immunohistochemical study. Annals of Anatomy, 218:205–212, 7 2018.

[15] Joseph M. Abdo, Nikolai A. Sopko, and Stephen M. Milner.

The applied anatomy of human

skin: A model for regeneration, 3 2020.

[16] Sarah A Mohamed and Rachel Hargest. Surgical anatomy of the skin, 2021.

49

[17] Eliana B. Souto, Joana F. Fangueiro, Ana R. Fernandes, Amanda Cano, Elena SanchezLopez, Maria L. Garcia, Patr´cia Severino, Maria O. Paganelli, Marco V. Chaud, and

Am´elia M. Silva. Physicochemical and biopharmaceutical aspects influencing skin permeation and role of sln and nlc for skin drug delivery, 2 2022.

[18] What is cancer?, national cancer institute. https://www.cancer.gov/about-cancer/

understanding/what-is-cancer, 2021.

[19] Mahsa Alidoust Saharkhiz Lahiji and Fatemeh Safari. Potential therapeutic effects of

hamscs secretome on panc1 pancreatic cancer cells through downregulation of sgk269, ecadherin, vimentin, and snail expression. Biologicals, 2022.

[20] Rachita Pandya, Kaitlyn Grace San Diego, Talha Shabbir, Arnav P. Modi, Justin Wang,

Joseph Dhahbi, and Sanford H. Barsky.

The cell of cancer origin provides the most reliable

roadmap to its diagnosis, prognosis (biology) and therapy. Medical Hypotheses, 157, 12

2021.

[21] Skin cancer 101, skin cancer

skin-cancer-information/, 2022.

foundation.

[22] Skin

cancer

(non-melanoma),

non-melanoma-skin-cancer/, 2020.

nhs.

https://www.skincancer.org/

https://www.nhs.uk/conditions/

[23] By Mayo Clinic Staff. Skin cancer, mayo clinic. https://www.mayoclinic.org/

diseases-conditions/skin-cancer/symptoms-causes/syc-20377605, 2020.

[24] Cancer, wikipidia. https://fr.wikipedia.org/wiki/M%C3%A9lanome, 2022.

[25] Squamous cell carcinoma, walk-in dermatology.

patient-education/squamous-cell-carcinoma/.

https://walkindermatology.com/

[26] First immunotherapy for advanced basal cell carcinoma approved in us, european pharmacautical review. https://www.europeanpharmaceuticalreview.com/news/142524/

libtayo-first-immunotherapy-for-advanced-basal-cell-carcinoma-approved-in-us/,

2021.

[27] Skin cancer prevention, skin cancer foundation.

skin-cancer-prevention/, 2022.

https://www.skincancer.org/

[28] By Mayo Clinic Staff. Skin cancer diagnosis, mayo clinic. https://www.mayoclinic.org/

diseases-conditions/skin-cancer/diagnosis-treatment/drc-20377608, 2020.

[29] Artificial intelligence, built in. https://builtin.com/artificial-intelligence.

[30] JAKE FRANKENFIELD. Turing test, investopedia. https://www.investopedia.com/

terms/t/turing-test.asp, 2022.

[31] machine learning, built in. https://builtin.com/machine-learning.

[32] Deep learning: An in-depth look into the ai-based tech, built in. https://builtin.com/

artificial-intelligence/deep-learning.

[33] What is computer vision?, ibm. https://www.ibm.com/topics/computer-vision.

[34] machine learning in computer vision, full scale.

machine-learning-computer-vision/, 2019.

50

https://fullscale.io/blog/

[35] M. Krishna Monika, N. Arun Vignesh, Ch Usha Kumari, M. N.V.S.S. Kumar, and E. Laxmi

Lydia.

Skin cancer detection and classification using machine learning.

volume 33, pages

4266–4270. Elsevier Ltd, 2020.

[36] Daniella Castro Ara´

ujo, Adriano Alonso Veloso, Renato Santos de Oliveira Filho,

Marie Noelle Giraud, Leandro Jos´e Raniero, Lydia Masako Ferreira, and Renata Andrade

Bitar. Finding reduced raman spectroscopy fingerprint of skin samples for melanoma diagnosis through machine learning. Artificial Intelligence in Medicine, 120, 10 2021.

[37] What is raman spectroscopy?, edinburgh instruments ltd. https://www.edinst.com/

blog/what-is-raman-spectroscopy/.

[38] Mohammad Ali Kadampur and Sulaiman Al Riyaee. Skin cancer detection: Applying

a deep learning based model driven architecture in the cloud

for classifying dermal cell

images. Informatics in Medicine Unlocked, 18, 1 2020.

[39] Andre G.C. Pacheco and Renato A. Krohling. The impact of patient clinical information

on automated skin cancer detection. Computers in Biology and Medicine, 116, 1 2020.

[40] Priyanti Paul Tumpa and Md Ahasan Kabir.

An artificial neural network based detection and classification of melanoma skin cancer using hybrid texture features.

Sensors

International, 2, 1 2021.

[41] Simon M. Thomas, James G. Lefevre, Glenn Baxter, and Nicholas A. Hamilton. Interpretable deep learning systems for multi-class segmentation and classification of nonmelanoma skin cancer. Medical Image Analysis, 68, 2 2021.

[42] Hennie de Harder.

Model-agnostic methods for interpreting any machine

learning

model,

towardsdatascience.

https://towardsdatascience.com/

model-agnostic-methods-for-interpreting-any-machine-learning-model-4f10787ef504,

2020.

[43] Deepak Painuli, Suyash Bhardwaj, and Utku k¨ose. Recent advancement in cancer diagnosis

using machine learning and deep learning techniques: A comprehensive review, 7 2022.

[44] By Mayo Clinic Staff.

Melanoma, mayo clinic.

https://www.mayoclinic.org/

diseases-conditions/melanoma/symptoms-causes/syc-20374884, 2022.

[45] Melanoma overview, skin cancer foundation.

skin-cancer-information/melanoma/, 2022.

https://www.skincancer.org/

[46] Yamashita Rikiya, Mizuho Nishio, Do Richard Kinh Gian, and Kaori Togashi.

Convolutional neural networks: an overview and application in radiology.

Insights Into Imaging,

9, 2018.

[47] Arief Budhiman, Suyanto Suyanto, and Anditya Arifianto. Melanoma Cancer Classification

Using ResNet with Data Augmentation. 2019.

51