T.C.

ANTALYA BILIM UNIVERSITY

INSTITUTE OF POSTGRADUATE EDUCATION

ELECTRICAL AND COMPUTER ENGINEERING

THESIS PROGRAM

DERMATOLOGIST-LEVEL CLASSIFICATION OF SKIN CANCER WITH DEEP NEURAL NETWORKS

DISSERTATION

Prepared By

Junaid IQBAL

ANTALYA – 2021

T.C.

ANTALYA BILIM UNIVERSITY

INSTITUTE OF POSTGRADUATE EDUCATION

ELECTRICAL AND COMPUTER ENGINEERING

THESIS PROGRAM

DERMATOLOGIST-LEVEL CLASSIFICATION OF SKIN CANCER WITH DEEP NEURAL NETWORKS

DISSERTATION

Prepared By

Junaid IQBAL

Dissertation Advisor

Asst. Prof. Dr. Shahram TAHERI

ANTALYA – 2021

APPROVAL/NOTIFICATION FORM
ANTALYA BİLİM UNIVERSİTY
INSTITUTE OF POST-GRADUATE EDUCATION

Junaid Iqbal a M.Sc. student of Antalya Bilim University, Institute of Post Graduate Education, Electrical, and Computer Engineering owning student ID 181222005, successfully defended the thesis/dissertation entitled "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Academic Tittle, Name-Surname, Signature

Jury Member (Chairman) **:** Prof. Dr. Shahram TAHERI…….

Jury Member **:** ……

Jury Member **:** ……..……...

Date of Submission: …….

Date of Defense: ……

Director of the Institute: ……

ÖZET

DERMATOLOG DERECESİNDE DERİN SİNİR AĞI İLE CİLT KANSERİNİN SINIFLANDIRILMASI

Son birkaç on yılda, cilt kanseri vakaları muazzam bir şekilde arttı. Hem gelişmiş hem de gelişmemiş ülkelerde bile erkek / kadın ölümlerinin en popüler nedenini alıyor. Cilt kanseri erken aşamalarda tedavi edilemezse, insan cildine inerek vücudun diğer bölgelerine hızla yayılabilir. Melanom, hastalara şans eseri görünür hale gelir ve erken aşamalarda tespit etme fırsatı yaratır. Bilgisayar destekli teşhis için derin öğrenme yöntemlerinin yakın zamanda ortaya çıkışı, insan uzmanların bir hastanın durumu hakkında bilinçli kararlar vermesini kolaylaştırabilecek akıllı sağlık hizmeti görüntüleme tabanlı teşhis sistemlerinin geliştirilmesine izin verdi. Bu raporda, cilt lezyonu sınıflandırması sorununa, özellikle deri kanserinin (melanom) erken evrelerde saptanmasına odaklanılacak ve bir deri lezyonu içeren dermoskopik görüntüyü kanserli veya iyi huylu. Bu çalışmayı tekrarlamak için hangi metodolojiyi ve parametreleri kullanacağımızı belirledik. Bu raporda, derin ağların ve evrişimli sinir ağlarının, farklı görüntü sınıflandırmalarında el yapımı özellik çıkarıcıların yerini nasıl aldığına ışık tutmaya çalıştık.

Anahtar Kelimeler: Melanom Algılama, Evrişimli Sinir Ağları, Derin Sinir Ağları, Görüntü Sınıflandırma, Bilgisayar Destekli Tanı Sistemleri

ABSTRACT

DERMATOLOGIST-LEVEL CLASSIFICATION OF SKIN CANCER WITH DEEP NEURAL NETWORKS

Over the past few decades, cases of skin cancer have increased enormously. It is getting most popular cause of deaths in men/woman's even in both developed and undeveloped countries. If skin cancer can't be treated in early stages, it can go down in human skin and spread quickly to other parts in the body. Melanoma causes luckily become visible to sufferers creating an opportunity to detect in early stages. The recent advent of deep learning methods for computer - aided diagnosis has allowed the development of intelligent healthcare imaging-based diagnostic systems that can facilitate the human experts in making informed decisions about a patient's condition. In this report will focus on the problem of skin lesion classification, specifically the detection of skin cancer (melanoma) in early stages, and present a deep learning (Convolution Neural Network) based approach to classify the dermoscopic image containing a skin lesion as cancerous or benign. We have identified what methodology and parameters to use to replicate this study. In this report we have tried to brought in light that how deep networks and convolutional neural networks are taking the place of handcrafted feature extractors in different image classifications.

Keywords: Melanoma Detection, Convolutional Neural Networks, Deep Neural Networks, Image Classification, Computer-Aided Diagnosis Systems

DEDICATION AND ACKNOWLEDGMENT

I extend my sincere gratitude to Prof. Dr. Shahram Taheri for his guidance, encouraging words and his support throughout this journey. I am delighted to work with them, and I deeply appreciate everything they've shown me. Their dedication to teaching, their students and their program is impressive and inspiring. I have learned and perfected under their guiding hand. I would especially like to thank all my professors who taught me and assisted me during this program.

I dedicate this thesis work to my parents, friends, teachers, family, oneself, and all my supporters who always believed in my abilities.

ACADEMIC DECLARATION

I hereby declare that this master's thesis titled " Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks" has been written by myself under the academic rules and ethical conduct of the Antalya Bilim University.

I also declare that the work attached to this declaration complies with the university requirements and is my work.

I also declare that all materials used in this thesis consist of the mentioned resources in the reference list. I verify all these with my honor.

… /…/ 2021

Junaid Iqbal

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF CHARTS

ABBREVIATIONS

ANN	Artificial Neural Networks
BPPC	Binary Pattern of Phase Congruence
BSIF	Binarized Statistical Image Features
CAD	Computer Aided Diagnosis System
CLBP	Complete Local Binary Patterns
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
GDP	Gradient-Based Patterns
GLTP	Generalized Local Ternary Patterns
HAM	Human Against Machines
HOG	Histogram of Gradients
ISIC	International Skin Imaging Collaboration
IWBC	Improved Weber local descriptor
LAP	Local Anisotropic Pattern
LBP	Local Binary Patterns
LDA	Linear Discriminate Analysis
LDiP	Local Directional Pattern
LDN (LDN)	Local Directional Number Pattern

LDTP	Local Directional Ternary Pattern
LFD	Local feature descriptor
LGBPHS	Local Gabor Binary Pattern Histogram Sequence
LGiP	Local Gradient Increasing Patterns
LGP	Local Gradient Patters
LTP	Local Ternary Pattern (LTP)
MBC	Monogenic Binary Coding
MBP	Monogenic Binary Patterns
MIB	Magnification Independent Binary Classification
MIM	Magnification Independent Multi Classification
ML	Machine Learning
MRELBP	Median Robust Extended Local Binary Patterns
MSB Classification	Magnification Specific Binary
MTP	Median Ternary Pattern
MVS	Multi-View Stereo
MWLD	Weber Local Descriptor
PAD	Presentation Attack Detection
PCA	Patient-Controlled Analgesia
POEM	Patterns of Oriented Edge Magnitudes

RBM	Small Boltzmann machines
ReLU	Rectified Linear Unit
RGB	Red Green Blue
RNN	Recurrent Neural Network
SFM	Structure-from Motion
SVM	Support Vector Machine
WHO	World Health Organization

CHAPTER ONE

1. Introduction

When a skin cell which is responsible for pigment production which are known as by the name of melanocytes grows uncontrollably then people get affected skin cancer or melanoma. By melanoma in men mostly effected areas are chest and back whereas in woman's get affected mainly on legs. Melanoma is the most dangerous type of skin cancer. [1] In some it has been also found that it can also appear on other parts of body like face and eyes. Some of the major risk factors of skin cancer are excessive sun exposure, a previous family history and fair skin. According to a CNN report there is an estimation of around 151,000 cases of melanoma each year by the year 2030 which is pretty shocking just above 96000 in 2019 if it continues as with current rate. The National Cancer Institute (NCI) reported that about 1 million Americans are living with melanoma [1].

There are two basic forms of skin cancer which are called benign and malignant [2]. Dermoscopy is a new visual inspection technique that magnifies the skin and removes reflective surfaces. Research found that the diagnostic accuracy with dermoscopy with proper training amounts to 75% -84%. [3] A CASH rule (Color, Design, Symmetry and Homogeneity) has been developed to improve scalability for dermoscopic expertise, procedural strategies such as 3-point check-lists, ABCDE rules (Asymmetry, Boundary, Color, Dermoscopic Structure and Evolving), and 7-point check-lists. Many doctors, however, miss these approaches to rely on their own experience [3].

The diagnostic accuracy of melanoma is estimated to be between 65 and 80 percent without a computer-based assistance. Diagnostic accuracy of skin lesions by 49 percent is improved with dermoscopic images [4].

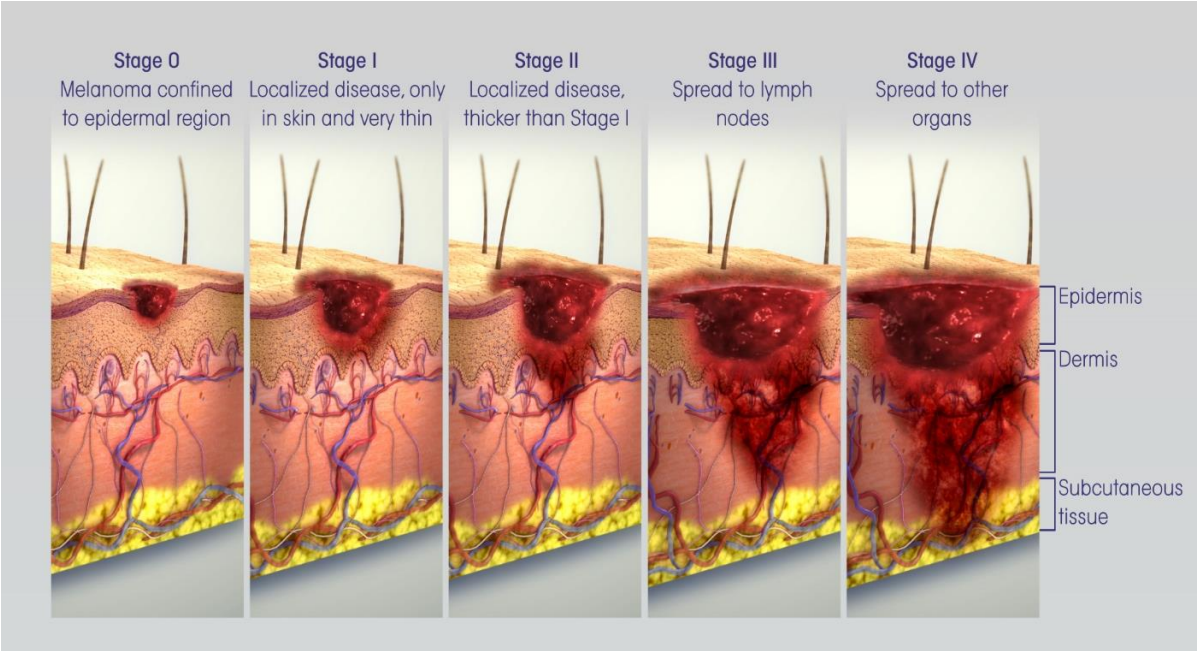

Figure 1: Skin Cancer Stages [5]

Skin cancer is staged in four parts, depending on the distribution of the disease. Stage 0 is often called melanoma in situ: the tumor cells are only in the outermost layer of skin (epidermis); level I and level II indicate that the cancer cells are in the layer immediately below the epidermis or touching down the next layer, yet there is no evidence that they have expanded to lymphatic vessels or other sections of the body.

As described, a medical imaging-based skin lesion and diagnosis system can be an appropriate tool to support a dermatologist in classifying skin cancer [4]. In this study, we are concerned in particular two class classification problems concerning melanoma and benign skin lesion.

Deep learning is dominantly getting used from image classification, face recognitions, tumors detection, and iris recognition to do different AI-related systems. From over the past 20 years enormous no of new techniques and algorithms are developed using deep learning techniques. Artificial intelligence and deep learning altogether have gained the attention of researchers for the last few years. Deep learning uses deep neural networks (DNN), a convolutional neural network with recurrent neural networks (RNN), and for supervised and un-supervised learning it uses Long-Short term memory.

The most efficient and powerful way to cope with complex classification and detection of different diseases is by deep learning classification algorithms. CNN (Convolutional Neural Network): is class of deep learning and applied for visualizing digital images. [6] CNN is based on multiple perceptron concepts because they are fully connected like input layers are connected to all hidden layers etc.

Machine learning phases involve the following four stages, collecting data, picking the suitable model, training the model and testing the model. Deep learning techniques have been used from so long time for skin cancer and other malignant. In our study we proposed different neural classifiers for detecting skin cancer mainly benign and malignant in early stages with more accurate results. Malignant tumors increase and spreads to nearby cells that can reach to other parts while benign tumors do not expand. Main purpose of this study is to develop effective machine learning approaches for skin cancer classification in a dataset by these different classifiers and to achieve better accuracy rate and results.

In this paper, we will illustrate that machine learning can save us from computationally time-intensive pre-processing steps, evaluate the performance of different lesion segmentation neural network models as an integral part of image processing-based techniques, and then compare neural networks for skin cancer detection on dermoscopic images using transfer learning and eventually evaluate machine learning for skin detection using transfer learning.

1.1. Contributions

We have provided an overview of the research contributions in this thesis towards the problem of skin cancer detection and classification at the early stages. At first, we provided the basic methods that are followed by researchers in their previous work, respectively what methods and techniques they used in their work. These steps and techniques helped in achieving some promising results in our respective work. Next, we described different CNN's and handcrafted features that are used for detection and classification. We thoroughly covered most of the CNN's and handcrafted feature descriptors for developing a better understanding related to the classification of skin cancer. We have investigated more than 30+ different architectures in both deep learning and hand-crafted domain and tried to find the optimal setting for each of the method. These architectures have been discovered by other authors for texture extraction or face recognition, facial expression recognition but never used in this domain, we believe they can be beneficial in this domain. Since previously achieved maximum accuracy was around 92%, we managed to obtain 89.89% accuracy by using Resnet101 on ISIC dataset. If implemented in real healthcare applications the increased accuracy means more patients can be diagnosed correctly, as this topic is directly related to skin cancer diagnosis, these better diagnoses will save more lives. We described a thorough study of the dataset including the pros and cons and also the problems that we faced using dataset. Lastly, we provide extensive evaluations to confirm the performance of the proposed methods compared to the state-of-the-art methods.

1.2. Thesis Organization

This thesis is organized in such a way that literature review and the previous researches and basic old methodologies are described in chapter-2. Chapter 3 contains post-applied approaches and also major approaches that we applied and can be applied in the future. Chapter 4 contains information on the dataset that we used for our work and chapter 5 consists of experiments that we performed using multiple approaches and our results that how they differ from other work. In this last chapter, we presented a conclusion about which approaches is better for what kind of image classifications.

CHAPTER TWO

2. Literature Review

Due to recent breakthroughs in computing capabilities and their relatively low cost, deep learning techniques are now becoming day by day more attractive to researchers [7]. In replacing the handcrafted features with a machine learning process, the results are able to converge much faster.

The ABCDE classification system will often be used by physicians to classify skin cancers. Some of the signs that are associated with the disease include asymmetry of the patient, the irregular borders of the area, the uneven color distribution in the area, or the size of the marks on the patient's body. While very exact, you will want to make sure your examples have the same color, same architecture, have the same number of points, are symmetrically designed, and look similar across all examples (CASH) [1]. Although many examinations can be subjective in nature, visual examinations are fairly easy to evaluate. Skin lesion analysis includes very challenging visual classification and segmentation tasks because of the wide inconsistency in the appearance of skin lesions due to moles and also because of the wide variance in the variability of skin lesions among different patients.

Neural network's idea is based on neurons that how they are interconnected with each other. Their working principle follows the working of biological neural networks. Dendrites are the input and output channels of biological neural networks. An ANN forms a highly connected network due to its millions of processing units which in results processes a large amount of information. A neural network is composed of 1 input layer and 1 or more hidden layer which is connected to the output layer. Simple deep neural network layers are defined in Fig. 1 that how each input image goes through different layers and how features and features maps are extracted from the input image. Prior in the research of skin cancer recognition, machine learning procedures were used vastly. For skin cancer detection, Situ N et al. used SVM [7]. For classification purposes, the Naïve Bayes classifier and KNN were also tried.

Researchers in machine learning and computer vision have attempted to classify melanoma and analyze general skin disease, and we evaluate some of these related past methodologies in this section.

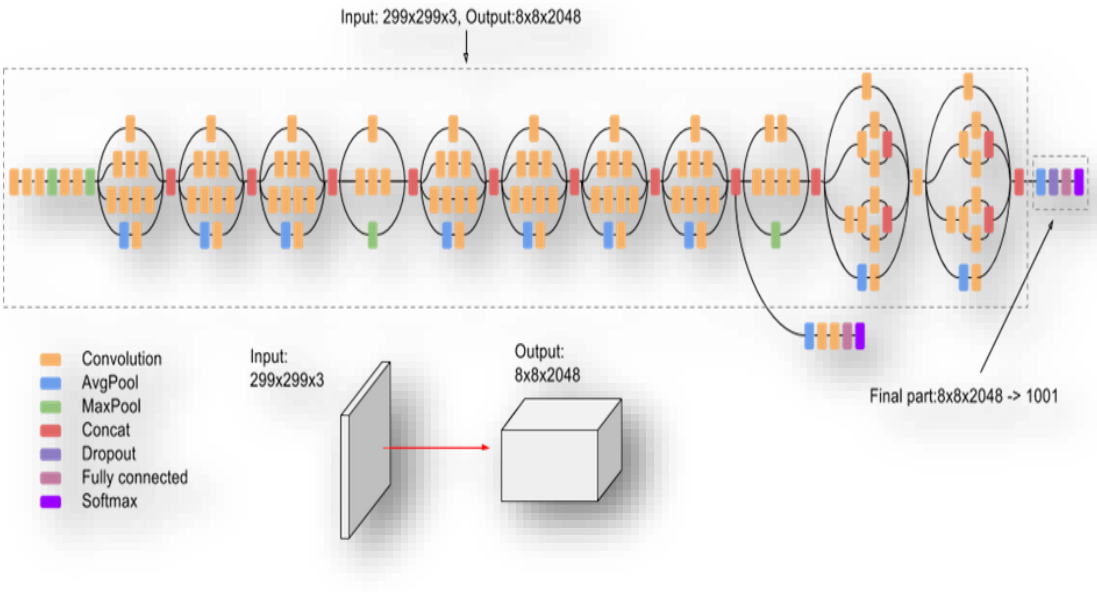

Figure 2: Deep Convolutional Neural Network InceptionV3 [8]

In this paper, researchers identified 2,032 different diseases arranged in a tree structure with three spurts representing individual classes: benign lesions, malignant lesions, and non-neoplastic lesions. It was derived by dermatologists using a top-down approach: individual diseases were assigned as vertices, then connected based on visual similarity, until the entire graph was connected. This aspect of the taxonomy is a useful way to create both appropriate training sets for machine learning models and ideal training sets for medical applications. The root nodes are the most fundamental classifications and are used as the first validation method. The children of the root nodes are used as the second validation strategy, and are combined into disease classes for which similar clinical treatments exist [9].

Computational intelligence techniques are used to assist both dermatologists and physicians in faster data processing to give better and more reliable diagnoses. This paper describes an automatic decision support system that utilizes melanoma images to identify characteristic objects commonly found in pigmented skin lesions. They present a review of

the current available technologies for skin imaging, and new research on noninvasive diagnosis of melanoma [10].

There are two basic types of skin cancer, non-melanoma and melanoma, much rarer, however a much more serious disease. In this paper, they proved that computerized adaptive neural networks can work exceptionally well in skin disease detection. Different from Agostinelli's method, this paper utilized a more productive adaptive activation function. This piecewise linear function is both piecewise linear and can represent both convex and non-convex functions of the input. The deep CNN for skin lesion classification was built by employing the activation function for the network [11].

This paper discusses a new method that indicates how skin lesion diseases can be predicted through deep learning by statistical algorithms. They have shown that a deep CNN for skin lesion classification can be constructed by adding the adaptive piecewise linear function, which allows for a notable performance improvement in our model without significantly increasing the number of learnable parameters [13].

In the paper published by the author, this paper proposes a deep learning approach for skin lesion classification. Unlike previous approaches that used hand-trained features, they use trained convolutional networks to extract features. Deep neural network can take images with only a few steps. These are very efficient methods to solve the classification issues.

In this reported study authors investigated the transfer learning capability of the features extracted from a pretrained CNN for the classification of skin lesion images. The experimental results show that those transfer features have fairly positive results when caring for skin mole characteristics. (The experimenters found out that the transferred features were more effective than other methods of acquiring mass data.) The net was tested on a particular problem in special data sets; however, it is possible to transfer the data to another issue like skin cancer, which shows a huge potential for new discoveries in the subject of skin cancer research [16].

The solution is implemented using neural network libraries and packages, which is achieved using Python 3.6. In addition to Keras, TensorFlow, H5PY, OpenCV and Scikit-

Learn. In order to improve classification performance, the initial segmentation of dermoscopy images must be performed first to classify them into different types of skin lesions. This helps to more specifically select the relevant features within the lesion regions without using the whole images. The clinical specialist received years of training in order to perform dermoscopy. However, our implementation of deep learning models for image analysis could aid in the elimination of the requirement for clinicians to perform manual segmentation tasks [17].

CNN, as a deep learning framework, is used in this paper for the automatic detection of melanoma. CNNs are taking advantage of a set of powerful convolve filters. They can examine the different structures in the input images. Thus, when using CNN, the input is the image itself and the network automatically extracts the appropriate aspects of the image. A complex computational method based on deep learning that used clinical images has been implemented in this paper. This system was capable of detecting cases of benign melanoma. We have been able to increase the accuracy of the system by sending images through an illumination correction that has increased the system's ability to discriminate [18].

Author's Name	Merits	Demerits	Datasets	Accuracy
R. Suganya et al	Improve the precision of the classification using the help vector machine.		Dermweb 320 images. 100: NoMSL, 220: MSLs.	Sensitivity: 95.4% Specificity: 89.3% Accuracy: 96.8%.
M.A. Farooq et al	Use 10 hidden layer ANN for classification.		DermIS and DermQuest. Both melanoma and nonmelanoma type image.	80%
D. Csabai et al	Uses an ultrasound image as a Addendum to the dermoscopic image. Both		Data set collected from Dermatooncology Department of Semmelweis University. Dataset size = 248 images, with 73	For the both classifiers achieved specificity of at least 19% at 100% sensitivity

	AdaBoost and SVM are used for classification.		melanoma, 130 BCC, and 45 nevi.	and an AUC of 84%.
Anurag Chawla et al	Uses feed forward neural network on discriminant feature extracted from 326 images.	Not able to achieve higher accuracy.	New York University, Department of Dermatology, and from Dr. Stoecker private practice.	Mean success rate when only considered melanoma images:86.0% when 60%/40% used as training/testing.
J Abdul Jaleel et al	Uses ANN to classify with a back propagation algorithm for instruction. Gray Level Matrix (GLCM) used as a function. ANN: 7 neurons at the input layer, 4 neurons at the hidden layer, 1 neuron at the output layer.		Image dataset collected from internal and hospital.	93%
Lequan Yu et al	Uses CNN for classification to exploit the image's local function.		Dataset of Skin Lesion Analysis Towards Melanoma Detection released with ISBI 2016.	Achieved accuracy of 94.9% on FCRN-50 network.

Table 1: Performance Comparison [19]

In this study, we applied deep learning techniques on our publicly available ISIC dataset by using 70% data for training and 30% data for testing using deep learning techniques. We used CNN classifiers and handcrafted feature descriptors for the classification of our data set. Our proposed classifiers achieved great accuracies with some précised results.

2.1. Deep learning Approach using Pre-trained neural network

Image classification is a particular process that takes some input as an image and gives output as a class. Image classification is challenging mainly microscopic images from the histopathologic section because they have issues like a large amounts of intra-class variability, rich geometric structures, and complex textures. Deep learning explores results from input data rather than hand-crafted features.

2.1.1. Deep Neural Networks

A deep neural network (DNN) is essentially an artificial neural network (ANN) with several layers between the input and output layers. The deep neural network seeks the right mathematical manipulation to transform input data to output, whether that is a linear relationship or a non-linear relationship. The network passes through the layers to determine the likelihood of each production. For instance, a DNN trained to identify dog breeds will go over the given image and measure the likelihood that the dog in the image is a certain breed.

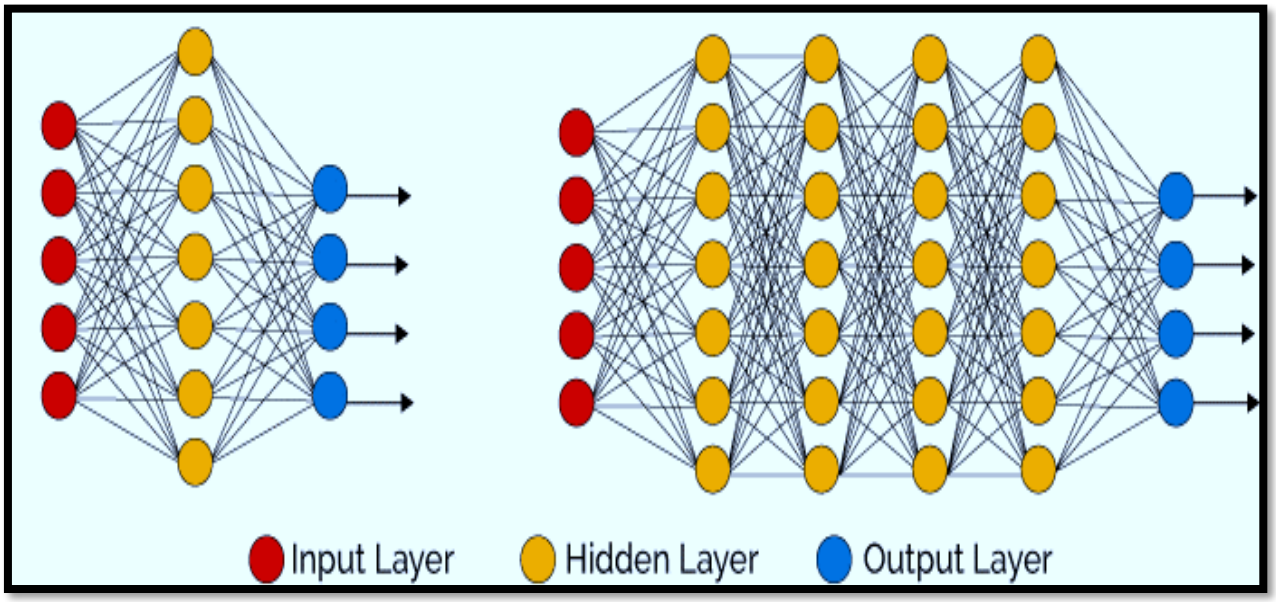

Figure 3: Simple Neural Network vs Deep Learning Neural Network [20]

DNN is also called Convolutional Networks. In DNN computations are performed by different layers. Computations for a network of the hidden layer is denoted as:

$$f(x) = f[a^{(L+1)} \ (h^{(L)} \ (a^{(L)} \ ](\ldots(h^2 \ (a^2 \ (h^1 h \ (a^1 \ (1) \ (x) \ ) \ ) \ ) \ ) \ ) \ ) \ [21]$$

Each pre-activation function a (l) (x) is typically a linear operation with matrix W (l) and bias b (l), which can be combined into a parameter θ:

$$a^{(l)}(x) = W^{(l)}x + b^{(l)} \text{ [21]}$$

$$a^{(l)}(x) = \theta^{(l)}x, 1 = 1 \text{ [21]}$$

$$a^{(l)}\, h^{(l-1)} = \theta^{(l)} h^{(l-1)} \quad \text{[21]}$$

Hidden-layer activation functions h(l)(x) often have the same form at each level, but this is not a requirement.

A deep neural network is the hierarchical architecture of neurons where each neuron is connected to other neurons [22]. Same as neuron this learning method has layers that pass messages from one layer to another and form a complex network that learns with some feedback mechanism.

A deep learning algorithm Convolutional neural network also called CovNet takes an input images and assign some sort of weights or biases to this image so that it can differentiate one from another. Covnet has very much less preprocessing as compared to other algorithms. [22] Filters are hand engineered in primitive methods, Covnets are used to avoid hand engineered algorithms because they are pretrained on datasets. Deep learning is comprised of numerous algorithms that follows hierarchal structures. Deep transfer learning is based on CNN. CNN's are pretrained networks and they classify ImageNet [22].

An algorithm learns the same way as a child learns because a lot of images are to be showed to algorithms for training and testing. Computers 'see' the world in a different way than we human beings do. Their world consists of only values and numbers. Every image represented as 2-dimensional arrays of numbers can be known as pixels. CNN Architecture is composed of three layers. Convolutional layer is composed of extracting features that extracts input features such as endpoints, corner, and edges. Convolution layer extracts features. Pooling Layer Function Reduces the dimensionality of each feature map but retains

the most important information. The pooling layer decreases the image resolution which reduces the precision accuracy of the effect (shift and distortion). All the activations in the previous layer are completely connected to the layer. They act as a classification system and produce the production. Fully connected, each two neurons are connected in each layer.

2.2. Proposed Methods

Our main purpose of this work is to make awareness about pre-trained neural networks for skin cancer classification. As we already know that machine learning is basically a branch of artificial intelligence, a science that explores machines by learning from existing dataset to get new information, new skills and new abilities and to distinguish existing information. Machine learning is mostly used in computer vision, data mining, strategy games, biometrics, robotics, DNA sequence sequencing, speech and handwriting recognition, detection of credit card fraud, medical diagnostics, securities market analysis, search engines & natural language processing.

Some of the pre-trained neural networks include Restnet201, InceptionV2 and V3, GoogleNet, densenet201, resnet-101, resnet-18, mobilenet-v2 etc. In our dataset we have two classes benign and malignant. According to theorems and old research work we know that feed forward network with single input is enough for representing any function. Since Alexnet have 5 convolutional layer and Google net have 22 layers still it is hard and challenging for training a deep network because of vanishing gradient problems and data over-fitting. [23] The main objective of introducing Resent and other pretrained neural network in deep learning world was to “identify shortcut connections” that can jump one or more layers. Similarly, resnet18 is another pretrained neural network that contains 18 layers, resnet-101 includes 101 layers, and similarly densenet201 contains 201 layers. Their main concern is to classify millions of objects with much accuracy and with high performance. Transfer learning techniques are widely in used these days. Transfer learning improves the accuracy of classifier and also accelerates the learning process.

Machine learning mainly is classified into following:

1. Unsupervised learning: Info information has no labels, yet calculations to gather the characteristic connections of information, for example, grouping and affiliation rule learning. Regular calculations incorporate autonomous segment examination, Apriority and K-Means calculations.

2. Supervised learning: Input information is labeled. Managed learning sets up a learning procedure, contrasts the anticipated outcomes and the genuine aftereffects of the "preparation information" (i.e., input information), and ceaselessly modifies the prescient model until the anticipated consequences of the model arrive at a normal exactness, for example, order and relapse issues. Normal calculations incorporate choice trees, Bayesian order, least squares relapse, strategic relapse, bolster vector machines, neural systems, etc. [23].

3. Semi-supervised learning: input information part labels, is augmentation of regulated learning, regularly utilized for characterization & relapse. Normal calculations incorporate diagram hypothesis deduction calculations, Laplacian bolster vector machines, etc.

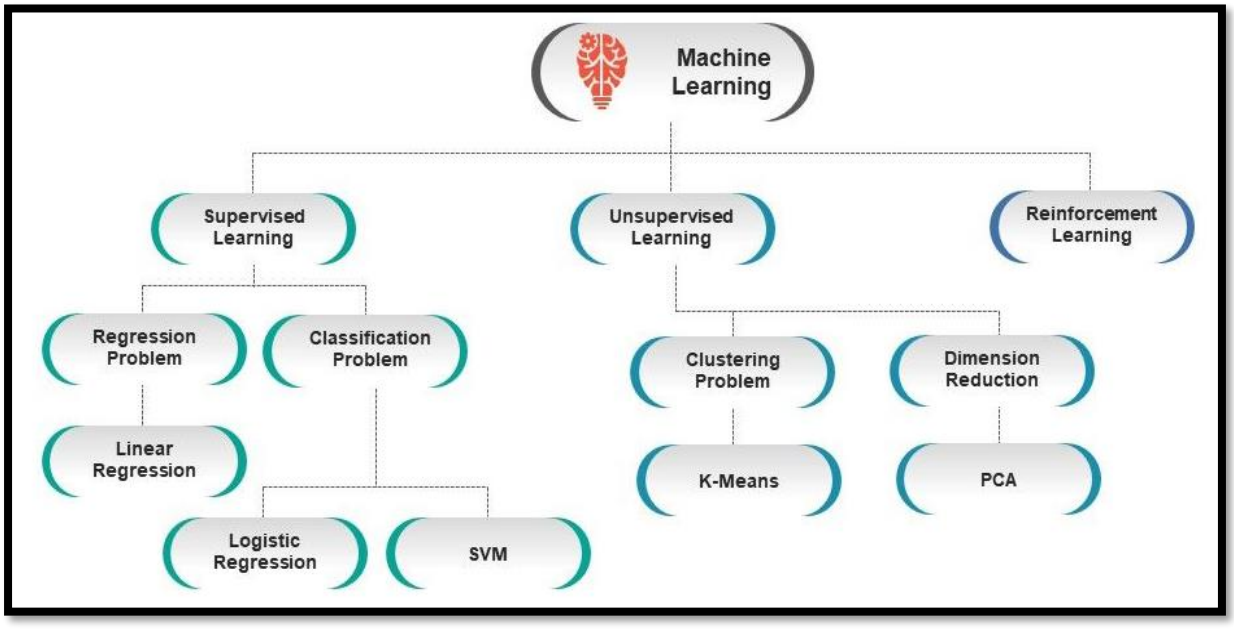

Figure 4: Supervised Learning vs Unsupervised Learning [24]

Reinforcement learning: Information as criticism to the model, underscoring the proper behavior dependent on the earth to expand the normal advantages. The distinction between regulated learning is that it doesn't require the right information/yield matches and doesn't require exact revision of imperfect conduct. Support learning is progressively centered on web-based arranging and requires a harmony between investigation (in the obscure) and consistence (existing information).

2.2.1. Hand-Crafted Features

Hand-crafted features are being used from past of the years for image classification and object detection. Most commonly hand-crafted features are buildup of 3 major techniques: Local Phase Quantization (LPQ), Local Ternary Pattern (LTP) and Local Binary Patterns (LBP) Histogram of Oriented Gradients (HOG) and Gabor filter. We have used 27 different hand-craft feature descriptors in our report. Mostly feature descriptors are based on LBP. [25] LBP feature extraction techniques is widely in used by researches for Computer vision technologies such as facial recognition and recognition of the face of expression, recognition of finger-veins, and estimation of human age.

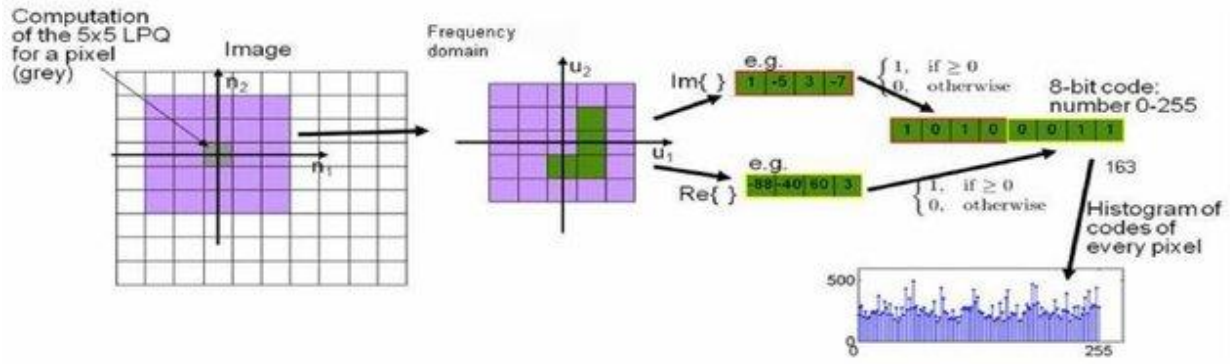

Figure 5: Workflow of LPQ feature extraction algorithm (left to right) [26]

With time number of features is increased and has shown some enormous results in object detection and image classifications. Feature based approaches are mainly combination of different features.

In different applications systems, biometric recognition systems are now commonly used because they are difficult to steal, are highly reliable to recognize and convenient for users. The disparity between people's physical and behavioral properties is based on this recognition process. For example, fingerprints and faces have been used historically for identifying a person. Several biometric features have been used in identification Face and fingerprint features such as iris, veins, ears, or pal-prints [27]. Methods for the face recognition program have been investigated for the presentation attack detection (PAD) [27]. Past researches are divided into two groups of methods of extraction based on non-training and training-based feature extraction. As the first level, the sparse low rank, bilinear, differential image model used by Tan et al. to distinguish images from other forms of presentation-attack, derived by varying Gaussian (DoG) [27] or logarithmic total (LTV) approaches. They showed that the real and presented image attacks can be discriminated against using its own tool using the NUAA database. Maatta et al. [27] used three feature extraction methods, Gabor filter, Local Phase Quantization (LPQ) and Local Binary Pattern (LBP), with the NUAA database, to extract image features and identify actual and presentation attack images using vector support machines (SVMs). The classification error was substantially reduced compared to those in Tan et al's study, based on their results.

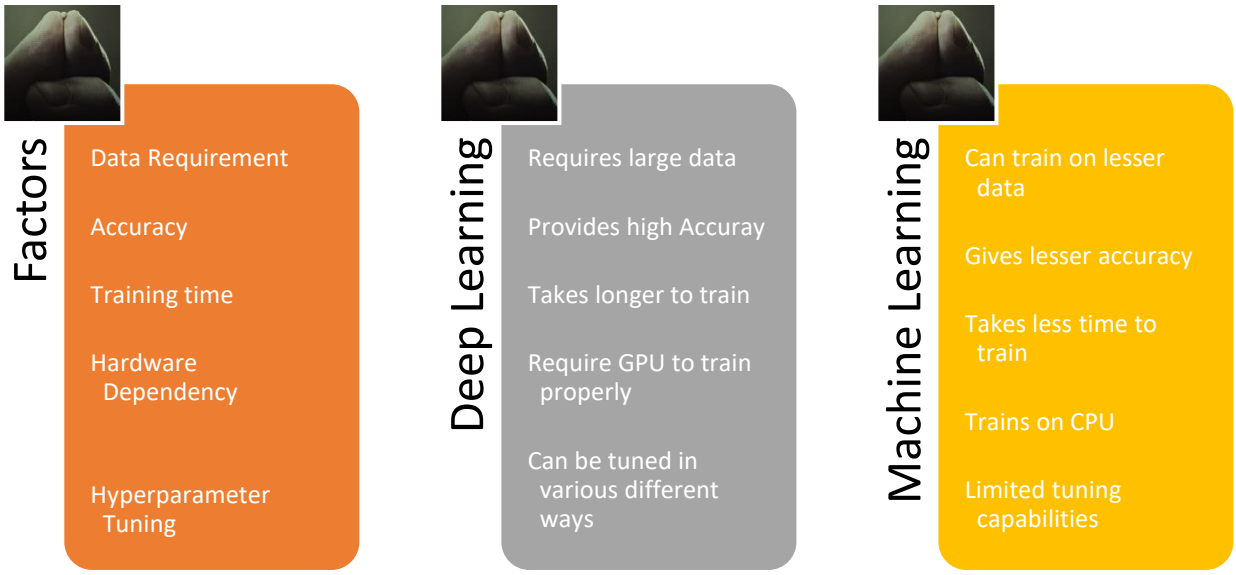

Figure 6: Machine learning vs. Deep Learning

The above figure includes complete difference between deep learning and machine learning that which one is easy to train or which gives more high accuracy etc. The number of features has increased over time to allow researchers to better adapt to the various tasks being dealt with. The features we have developed are based on LBP in our Local Binary Pattern, Combined Local Binary Pattern (CLBP) research. Both these approaches remove the overall complexity, so that a standard computer vision architecture can be better created.

2.2.2. Deep Learning-based Approaches

Pre trained architectures made to be used on different tasks, they select optimal setting according to the provided task, and these are included as transfer learning as they have already been trained and ready to perform on desired task. Deep neural networks mainly

called as artificial neural networks have multiple layers between input and output layers. They find precise mathematical manipulation in both linear and non-linear relationships and converts input layers to output. DNN is also called Convolutional Networks. From past 10 years, deep neural learning methods have become increasingly common because of the advanced technologies in language recognition, computer vision and other fields.

Our recent success can be attributed to increased data availability, improved hardware and software, and numerous algorithmic breakdowns, which accelerate training and generalize new data. Mostly used learning algorithms for classification and edge detection are shallow architectures (1-3 levels) (KNN, MOG, SVM, KDE, Parzen Kernel regression, perceptron, PCA)

DNN is also called Convolutional Networks. In DNN computations are performed by different layers. Computations for a network of the hidden layer is denoted as:

$$f(x) = f[a^{(L+1)} \ (h^{(L)} \ (a^{(L)} \ ](\dots(h^2 \ (a^2 \ (h^1 h \ (a^1 \ (1) \ (x) \ ) \ ) \ ) \ )$$

Each pre-activation function a (l) (x) is typically a linear operation with matrix W (l) and bias b (l), which can be combined into a parameter θ:

$$a^{(l)}(x) = W^{(l)}x + b^{(l)} \text{ [28]}$$

$$a^{(l)}(x) = \theta^{(l)}x, 1 = 1 \text{ [28]}$$

$$a^{(l)} \ h^{(l-1)} \ = \theta^{(l)} h^{(l-1)} \text{ [28]}$$

Hidden-layer activation functions h(l)(x) often have the same form at each level, but this is not a requirement.

A deep neural network is the hierarchical architecture of neurons where each neuron is connected to other neurons. Same as neuron this learning method has layers that pass messages from one layer to another and form a complex network that learns with some feedback mechanism. Fig 2 illustrates the input, hidden, and final layers of DNN.

A deep learning algorithm Convolutional neural network also called CovNet takes an input images and assign some sort of weights or biases to this image so that it can differentiate one from another. Covnet has very much less preprocessing as compared to other algorithms. Filters are hand engineered in primitive methods, Covnets are used to avoid hand engineered algorithms because they are pretrained on datasets [29]. Deep learning is comprised of numerous algorithms that follows hierarchal structures. Deep transfer learning is based on CNN. CNN's are pretrained networks and they classify ImageNet.

Covnet has very much less preprocessing as compared to other algorithms. While in primitive methods filters are hand-engineered, [29]. Figure down below illustrates how an input image goes through different stages in traditional machine learning while in deep learning algorithms all these features are read at single stage.

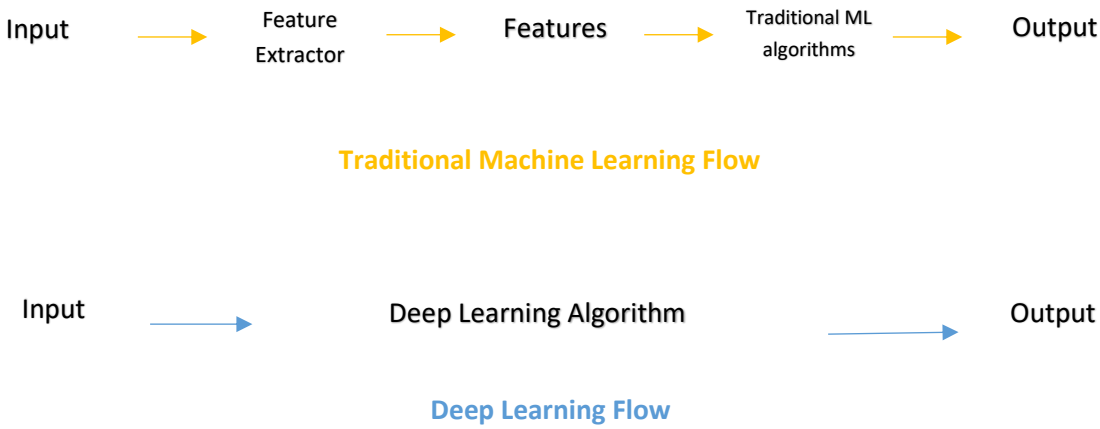

Figure 7: Machine learning flow vs. Deep learning Flow

Typically, in deep learning we have pre-trained architectures so normally we don't have to follow different parameters while in machine learning we have to create different parameter for better results. Researchers reported positive experiment results at normal levels of two or three, but the more profound training resulted in worse results. Scientists have been reporting for decades of deep neural networks.

Deep learning is a new process, which is used more and more in CAD systems and overcomes conventional computer education. Features are usually manually selected during computer education and fully automated in profound learning. Features like textures,

colors, and edges can be retrieved during machine learning, while other hierarchical or compositional features can be accessed during the training. Deep supervised network takes too much time for training and also, they are difficult to train [29]. CNN is less connected and less educated and have less parameters than deep supervised networks and it is easy to use.

The Table: 2 below shows the available pre-trained networks trained in ImageNet Dataset and some of their attributes. The network depth is known as the largest number of sequentially convoluted or completely connected layers on the path from the input layer to the output layer. RGB images are inputs to all networks.

Network	Depth	Size	Parameters (Millions)	Image Input Size
SqueezeNet	18	5.2 MB	1.24	227-by-227
GoogleNet	22	27 MB	7.0	224-by-224
Inceptionv3	48	89 MB	23.9	299-by-299
Densenet201	201	77 MB	20.0	224-by-224
Mobilenetv2	53	13 MB	3.5	224-by-224
Resnet18	18	44 MB	11.7	224-by-224
Resnet50	50	96 MB	25.6	224-by-224
Resnet101	101	167 MB	44.6	224-by-224
Xception	71	85 MB	22.9	299-by-299
Inceptionresnetv2	164	209 MB	55.9	299-by-299
Shufflenet	50	5.4 MB	1.4	224-by-224
Nasnetmobile	*	20 MB	5.3	224-by-224
Nasnetlarge	*	332 MB	88.9	331-by-331
Darknet19	19	78 MB	20.8	256-by-256
Darknet53	53	155 MB	41.6	256-by-256
Efficientnetb0	82	20 MB	5.3	224-by-224
Alexnet	8	227 MB	61.0	227-by-227
Vgg16	16	515 MB	138	224-by-224
Vgg19	19	535 MB	144	224-by-224

Table 2: Pretrained Networks [30]

CHAPTER THREE

3. DEEP LEARNING METHODS

3.1. Convolutional Neural Network (CNN)

For last few decades CNN has attained the attention of researchers because they have less parameters and easily trainable as compared to DNN. CNN are the most successful classifiers and tools for mostly accurate classification of skin cancers classification. Deep learning methods are usually classifiable into four categories: CNN-based techniques, small Boltzmann machines (RBMs). The CNN-based approaches have recently gained more and more attention around the world, producing promising results in literature [47]. A typical CNN framework consists of three types of layer: Convolution layer, pooling layer and fully connected layer. Complete CNN architecture is illustrated in figure listed below with all respective layers:

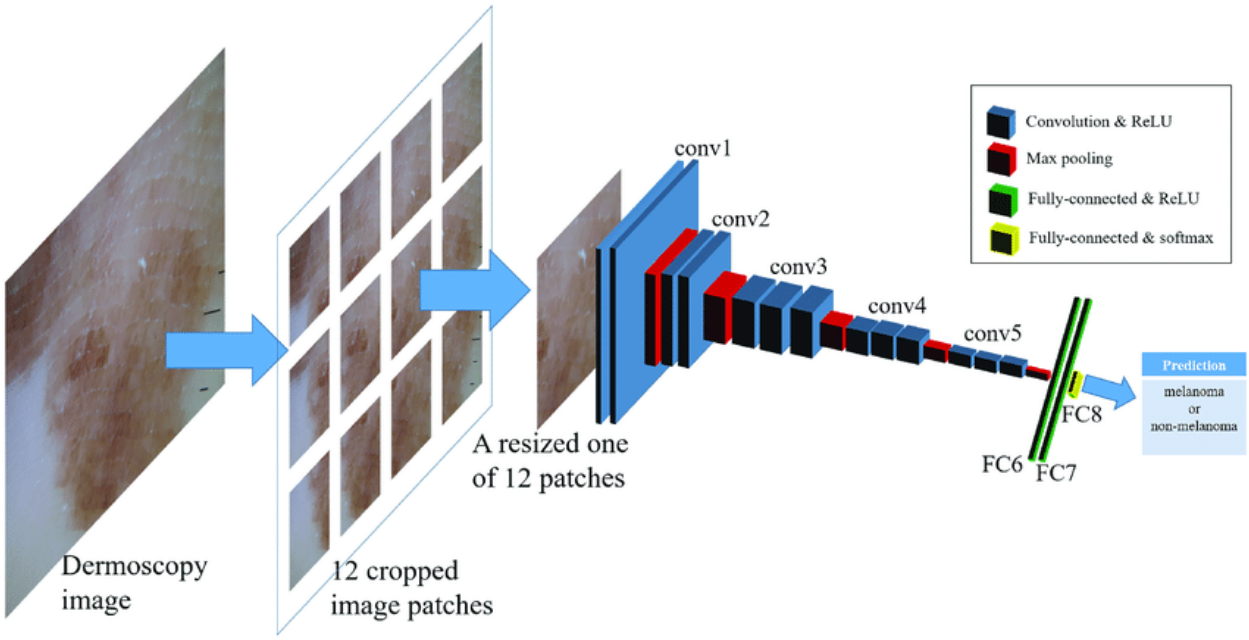

Figure 8: Schematic overview of CNN architecture for melanoma detection [48]

Discussing about principles of CNN, when an image is given to a CNN it tries in every possible position to match the required features. We make a filter by calculating match across the whole image and this math work done or obtained filter is convolution. Pooling layer extracts important information from large images and shrink images into small size after obtaining important information. A pooling layer is simply a pooling of an image or a series of images. The output has equal numbers of images, but each pixel is smaller. Taking an 8-megapixel image down to a 2-megapixel image makes it downstream much simpler for life [47]. Rectified Linear Unit or Relu is small but most important layer in CNN's.

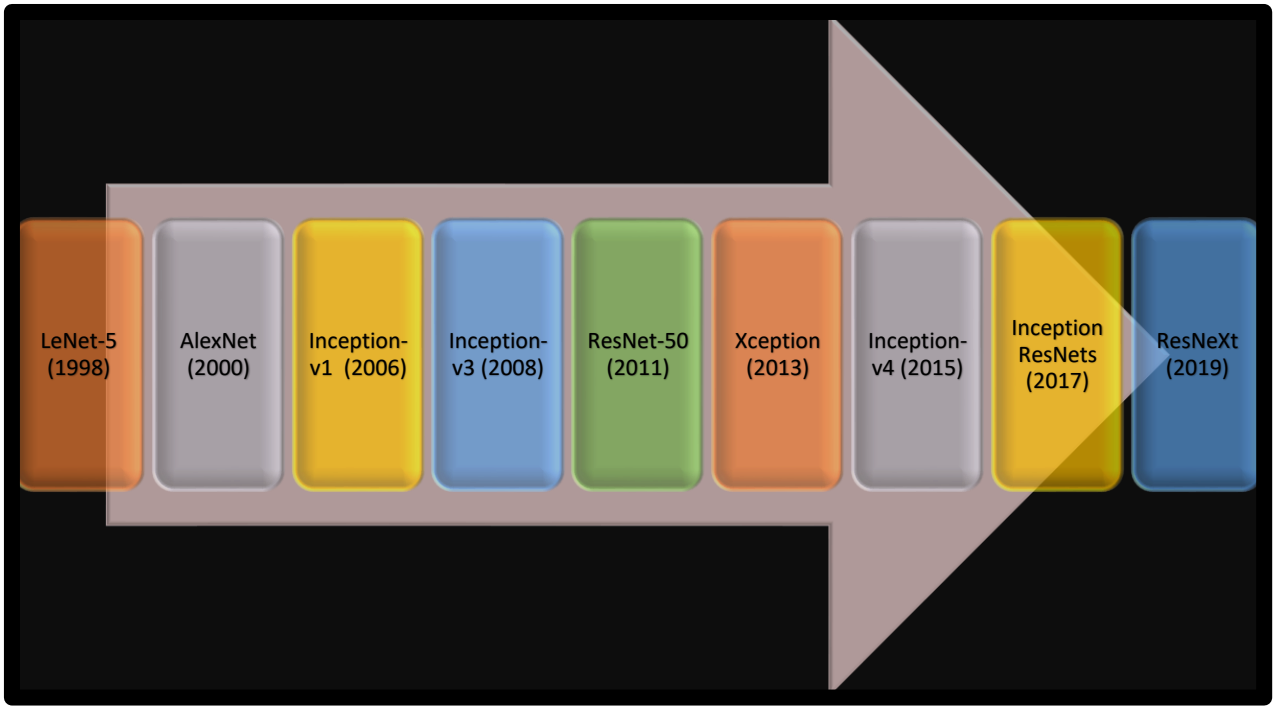

Figure 9: CNN architectures over a time from 1998-2019

The above explains the formation of various CNN architectures over time. The use of CNN-based imaging methods is distinct from natural imaging methods. On the other hand, a large labeled dataset, such as ImageNet, is required for training and testing CNNs. Rather than having RGB channels, medical images are usually grayscale. However, large-scale medical image databases are not always accessible due to comprehensive labeling analysis and the need for specialist knowledge [47]

3.2. Neural Network

Over the past 10 years, a methodology that is called "deep learning", has succeeded in the most powerful artificial intelligence systems. For example, speech sensors on smartphones, or Google's new automated translators. Deep learning is currently a modern concept for a neural networks technique that has been used and not used for more than 70 years. In 1944, two researchers at Chicago's McCullough and Walter Pitts proposed neural networks and, in 1952, moved to MIT, the director of the most widely recognized Cognitive Science Department. Neural networks were a key research field in both the neuroscience and the informatics until 1969, when the MIT mathematician Marvin Minsky and Seymour Papert, co-directing the new MIT Artificial Intelligence Laboratory a year later, destroyed them, according to computer science reports [49]. The strategy then made a comeback in the 1980s, was again eclipsed in the first decade of the new century, and returned, largely driven by the rising computing power in the graphical chips. A simple neural network with 2 input layered and several hidden layers is shown in figure below.

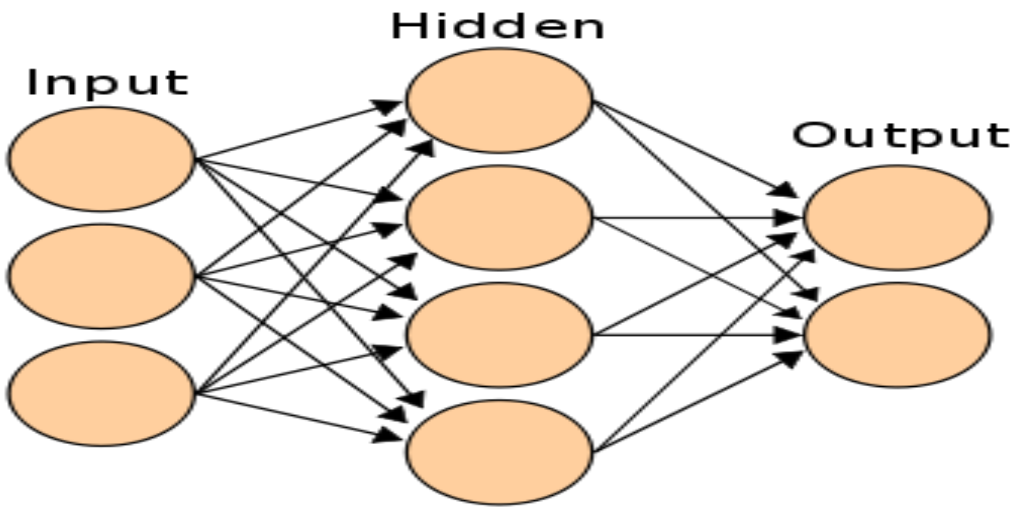

Figure 10: Simple Neural Network [49]

A simple Neural Network with input layers, hidden layers and output layers is illustrated in figure illustrated above. Neural networks are a way to teach machine learning, by analyzing training examples, a computer may do some research. The examples have typically been labelled manually beforehand. For example, an object recognition system may be used in the pictures, which consistently correspond with particular labels, by feeding thousands of

marked pictures of vehicles, homes and coffee cups, and so on. Incoming connections will be assigned a number known as a "weight" by nodes. After a neural net has been formed, all its weights and thresholds are set on a random basis [49]. The training data is transmitted to the lower levels – the input layer – and the following levels are reached, multiplying and adding together in dynamic ways before it eventually enters the output layer, dramatically transformed. The threshold and weights are continuously modified during preparation, until preparation data with the same labels reliably produce identical outputs. Some of the most popular pre-trained neural networks include GoogleNet, densenet201, resnet-101, resnet-18, Mobilenetv2 etc.

3.2.1. Resent

Resnet-1o1 is a pre-trained neural network and it consists of 101 layers. This architecture can be loaded on more than million images and provides outstanding results with great accuracy. Resnet-101 can classify images into huge no of objects e.g., dogs, face, gender, cancer etc. It takes an image input of 224-by-224.Similarly resnet18 is another pre-trained neural network that contains 18 layers and takes an input size image of 224-by-224 [50]. It I almost similar to resnet-101 as it also takes millions of images and has the ability to classify more than 1000 objects. Due to the notorious problem of vanishing gradient, deep networks are difficult to train. Resent is one of the first neural network classifier that skips one more layer for overcoming vanishing gradient problem.

Experiments show, however, that the Highway Network is nothing better than Resnet, a strange thing, given that the Highway network solution area contains Resnet and should therefore perform at least as well as Resnet [50]. This means that such "gradient highways" are more critical than a broader variety of solutions.

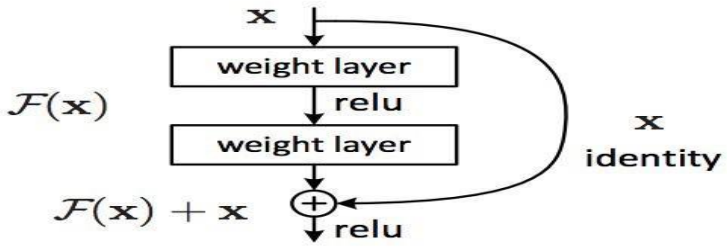

Figure 11: Residual Learning block [51]

Compared to a "simple" network in which new and distinct feature maps for each layer need to be learned if there is no clearing, the intermediate layers should then gradually change their weights to zero such that the residual block becomes an identity function as shown in above given figure.

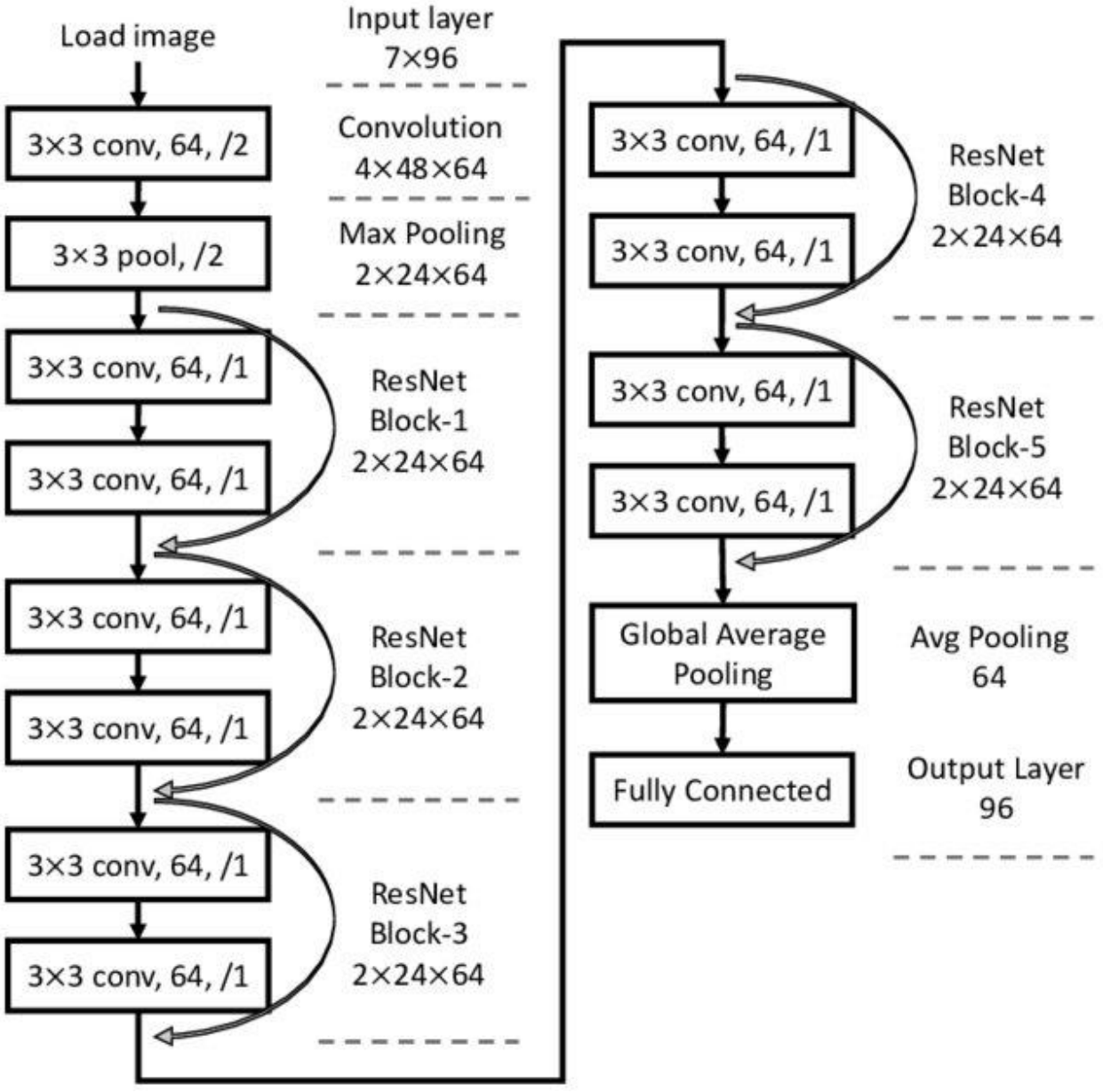

Figure 12: Resent-12 Structure [50]

Resnet is mostly in study and in attention of different researchers these days and few new variants of Resnet are also introduced in few years. Resnext is the new variant of Resent that is proposed by Xie at el [50].

Resnet's main advantage is that hundreds, even thousands of these residual layers can be used to create a network, and trained afterward. That is a little different from ordinary sequential networks, where you find the output enhancements are decreased as you increase the number of layers. For normal feedforward networks, the prediction accuracy decreases as the depth of the network increases. Many variables are responsible for each result, including vanishing of gradient problem, saturation, training data size, and over-fitting. Residual learning allows more than a thousand layers of network depth to become as deep as that. Skip connections make gradient flow easy during the backward pass, and solve the vanishing gradient problem.

3.2.2. Inception V3 (and V2)

Christian and his fellows are extremely professional scholars. Batch-normalized Inception was introduced in February 2015 as Inception V2. Batch-normalization measures the mean and standard deviation at the output of a layer of all feature maps. It involves "blanketing" the test, such that all neural maps have reactions in the same range and a mean of zero. It helps us to learn as the next layer no need to identify offsets in the data and can focus on integrate the best features should be integrated [53].

They announced a new version of the Inception modules in December 2015 and this article provides more information on the implementation choices and illustrates the original GoogleNet architecture further for the respective architecture. The original concepts are categorized into:

Maximize network information flow through the careful creation of networks that balance depth and width. Increase maps before any pooling.

The layer width or the number of features is also increased by increasing the depth.

To increase the feature combination to the next layer, using the difference in width on each layer.

Use only 3×3 convolution whereas 5×5 and 7×7 filters can be decomposed with multiples 3×3 filters as shown in figure below [53].

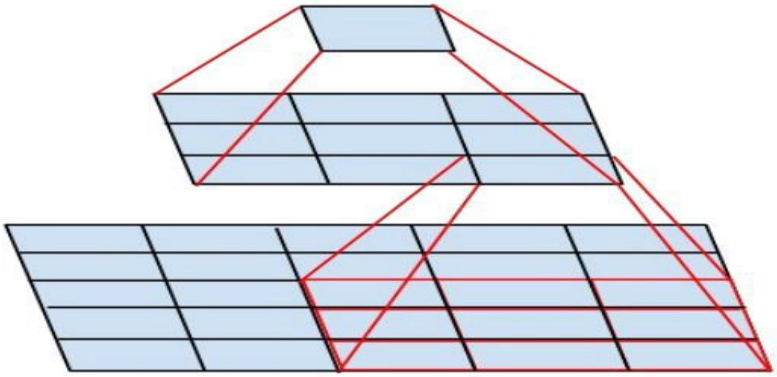

Figure 13: Decomposition of 5×5 filters into multiples 3×3 filters [52]

This architecture is comprised of total of 22 layers. Compared to previous Alexnet, ZFnet, and vggnet, it is already a very deep model. So, we can see that multiple initial modules are wired together to go deeper. (But not too deep in contrast to resnet created afterwards.) (There are some intermediate branches of SoftMax in the middle. SoftMax branches consist of 5x5 pooling layers having 3 stride and 1x1 convolutional with 128 filters with 1024 fully connected layers.

3.2.3. Densenet

Densenet also known as shortcut connection architecture—was proposed by Huang et al, this structure connects all the layers directly with each other. One of its well-known variants. Densenet-201 is also a renowned pre-trained neural network with having 201 deep layers. It also takes an input image of size 224-by-224 [54]. It can also classify millions of objects with much accuracy and with high performance. Each layer's input consists of feature maps of each earlier layer, which are then transferred to each layer. The characteristics of the maps are applied to depth. About this architecture authors say that it ensures feature reuse and makes the network highly parameter-efficient. One easy understanding is that the identity

mapping output has been applied to the next block, which may prevent information flow if the two-layer function map has very different distributions. Concatenated feature maps can therefore retain them all and increase the performance variance, which promotes reuse of the features. Densenet architecture with its layers architecture in listed down:

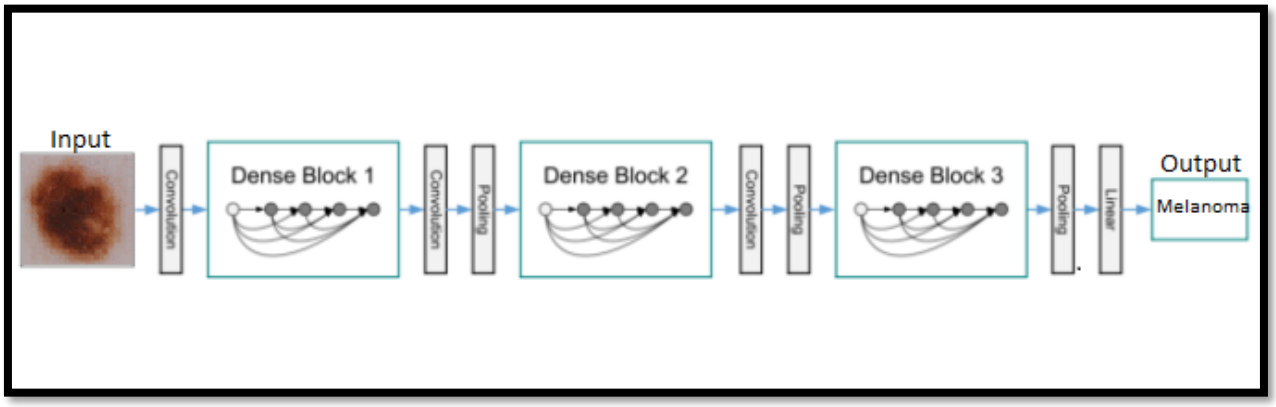

Figure 14: Layers Structure of Densenet [55]

Fig. 20 illustrates Densenet architecture that how can image moves through different layers and block and how features are extracted through these layers and at the end output is predicted in reference to extracted features. Due to this model, we know that all the layer is to have input maps of k * (l-1) + k 0 where the number of input image channels is k 0. The researcher used a growth rate(k) of hyper parameter and a 1x1 convolution bottleneck layer to reduce the number of maps until the expensive convolution (3x 3) is reached, to prevent the network from becoming too big. The complete layer to layer architecture is reflected below:

Layers	Output Size	DenseNet-121($k = 32$)	DenseNet-169($k = 32$)	DenseNet-201($k = 32$)	DenseNet-161($k = 48$)
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	[1 × 1 conv, 3 × 3 conv] × 24	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 48	[1 × 1 conv, 3 × 3 conv] × 36
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	[1 × 1 conv, 3 × 3 conv] × 16	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 24
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

Figure 15: Complete Flow Densenet Architecture [56]

3.2.4. Mobilenet

Mobilenet consists of 2 layers. The first layer is referred to as a depth wise convolution, and makes lightweight filtering with a single coevolutionary filter. Point-wise convolution which is second layer of mobilenet is 1x1 convolution, that is responsible for compiling new features by computing linear input channel combinations. Mobilenet_v2 consists of two different types of blocks. 1 block for drowsing and other residual blocks with stride 1. All forms of blocks have 3 layers. The first layer this time is 1x1 convolution with ReLU6 [57]. Depth-wise convolution is second layer. The third layer is another 1x1l, with no non-linearity. It is stated that if Relu is reused the deep networks would only have the power of a linear classifier on the output domain's non-zero volume portion. MobileNetV2 architecture with different layers and parameters in shown in Fig. 21.

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Figure 16: Mobilenetv2 architecture [57]

3.3. Transfer learning

The basic principle of transferring learning is to use information gained from activities for which there are many labeled data available in settings where there is only a limited amount of labeled data. It is expensive to construct labeled data, therefore making efficient use of existing datasets is important. The primary goal in a traditional model of machine learning is to generalize unknown data based on patterns learned from training data. With transfer learning, you seek to kick-start this cycle of generalization by starting from patterns learned for a specific task. Essentially, instead of beginning the learning cycle from a blank sheet (often initialized at random), you start from patterns that were learned to solve a particular problem [58]. In any form of learning transferring learning is important. To be successful in this, humans are not taught every single task or question. Everybody gets into challenges that have never been experienced, and we try to solve problems in an ad-hoc way. The opportunity to learn from a wide variety of perspectives and export 'information' to new worlds is just what learning transfer is all about. In that point, on a logical level, transfer learning and generalization are very close. The key difference is that learning transfer is mostly used to 'transfer information through tasks [58], rather than generalizing within a single task.

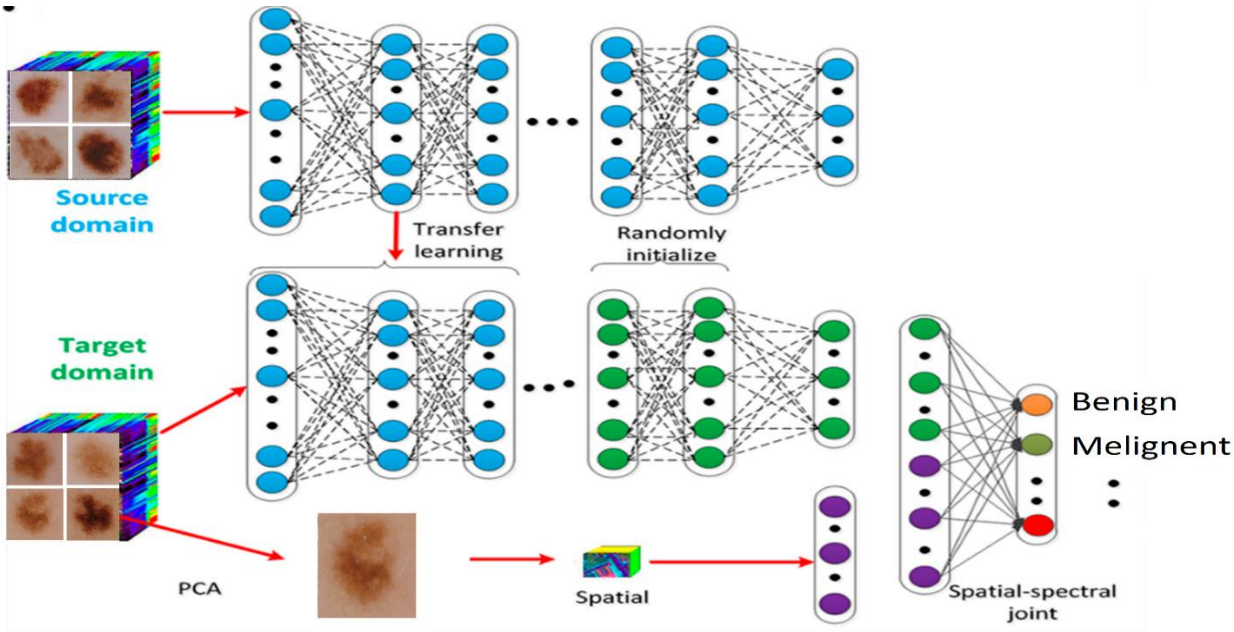

Figure 17: Transfer Learning Algorithm [59]

Figure above ensure the complete workflow of Transfer learning that how a new data is classified on fine-tuned pre-trained neural networks for better results. To ensure the advancement of deep learning techniques in a large number of small-data environments, transfer learning is crucial. Deep learning in science is pretty much everywhere, but many real-life situations usually don't have millions of data points labeled for training a model. Deep learning techniques require large data volumes to change the millions of parameters in a neural network. This means you need a lot of (highly expensive) labeled data, particularly in the case of supervised learning. [58] The labeling of images sounds simple, but expert expertise is needed in Natural Language Processing (NLP), for example, to construct a large labeled dataset. Transfer learning is one way to reduce the necessary size of datasets to provide a viable alternative for neural networks. Other viable options shift towards more probabilistically based models, which are usually best suited for handling small data sets.

As the name suggests, transfer learning involves the ability to transfer information from one domain to another. Transfer learning can be understood at a high level, i.e., NLP model architectures can be reused in sequence prediction issues as many NLP issues can be implicitly reduced to sequence prediction issues. Transfer learning can also be viewed at a low level, where parameters are simply imported from one model into another model (skip-gram, continuous bag-of-words etc.). Transfer learning techniques are widely in used these days. Transfer learning improves the accuracy of classifier and also accelerates the learning process.

CHAPTER FOUR

4. Dataset

4.1. ISIC (International Skin Imaging Collaboration) Dataset

The (ISIC) is a collaborative project that is meant to help improve skin cancer treatment through the use of skin imaging technologies. The International Society for Digital Imaging of the Skin is the organization behind this initiative (ISDIS). The ISIC Repository does contain an extremely large library of dermoscopic images that have been rated by certified dermatologists and can validly identify lesions.

While the ISIC (International Skin Imaging Collaboration) Database has over 13,000 dermoscopic images, it has been the result of efforts of acquiring from hundreds of medical centers stemming from different countries and has been acquired from various devices in each center [61]. It is broad and international participation that is designed to ensure that the sample of medically relevant images are clinically relevant.

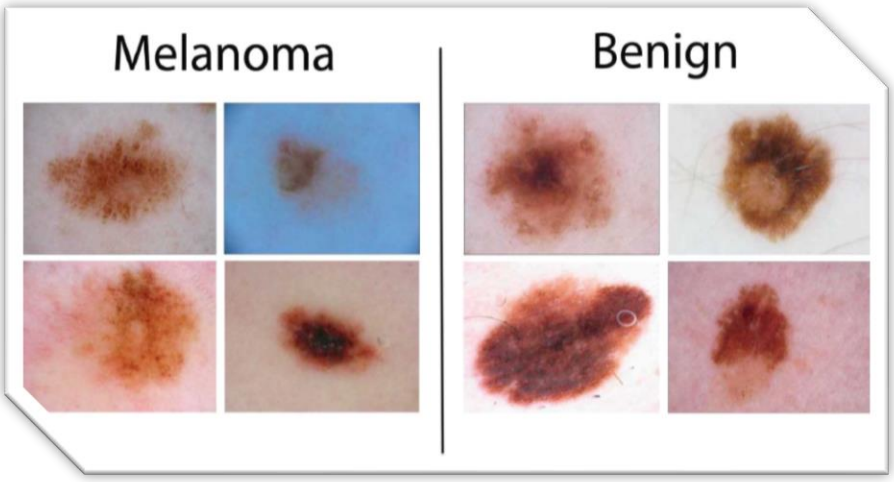

Figure 18: Sample images from the ISIC archive dataset [60]

For privacy and quality control, all incoming pictures in the ISIC archive are screened. In most images, clinical metadata have been associated and analyzed by renowned skin cancer experts. A subset of the samples is labelled and marked by well-reputed professionals on skin cancer. Note the dermoscopic features in the image, including

the international and focal morphologic components known to distinguish between skin lesions.

The ISIC Melanoma Project has as its overall objective to encourage the reduction of melanoma deaths and unnecessary biopsies by significantly improving skin cancer early detection's precision and effectiveness. In order to help other physicians, become familiar with the dermoscopic images, the ISIC is working on developing the dermoscopic images standards and writing a public archive of the skin lesions.

It is essential to improve melanoma classification for ensuring accuracy. The individual and economic costs of not screening for skin cancer are high. Screening procedures that are not thorough enough can lead to many unnecessary and inappropriate biopsies and skin cancer.

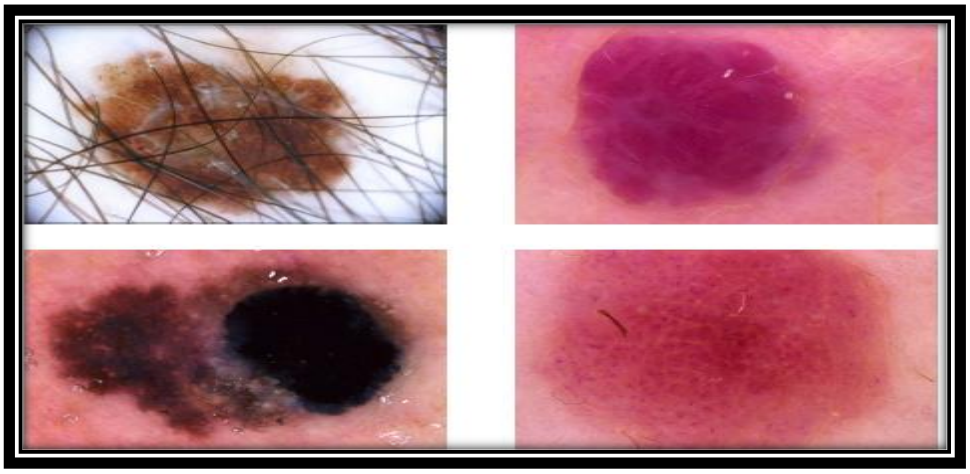

Figure 19: Sample of skin diseases from ISIC dataset [61]

The International Skin Imaging Collaboration (ISIC) started a repository of dermoscopic images. A base of initial research, which will eventually result in full automation of computer algorithms, of the dermoscopic images will be a big help to clinical training as well as research in the field of dermoscopy images. The ISIC holds an annual international competition where students create solutions to benefit users in the near future.

This dataset provides a balanced catalogue of pictures of benign skin moles and malignant moles. There are two folders, each containing 1800 images(224x244), of the mainly two type of skin lesion benign and malignant. Apart from pictures, the dataset also provides metadata for these pictures. It provides the patient's age and gender along with the location of the area affected by the disease.

CHAPTER FIVE

5. Hand-crafted Feature Descriptors

The ability of neural networks to learn data features higher than previous hand-made representations has recently brought significant changes in computer vision, e.g., recognition and object detection. Therefore, neural networks were implemented in order to create more inclusive representations of local features to the learning question of descriptors. The findings show major changes over conventional handcrafted representations such as SURF, DAISY or SIFT. In Fig. 8 a whole hand engineered architecture is presented that how an input image is classified and feature maps are extracted though different parameter, then presented to different feature descriptors which further classify for better output prediction.

Matching local image descriptors is an essential move for several computer-vision applications. This function has been used for more than a decade by hand-made descriptors such as SIFT. Several new descriptors obtained from the data were recently introduced and have shown that SIFT is more discriminatory. Matching the local image function is a significant step for many computer vision applications, including image recovery and picture-based localization for Structure-from Motion (SFM) and Multi-View Stereo (MVS) [31]. The overall output in many of these applications is highly dependent upon the quality of the initial matching feature level. A significant part of the computer vision community therefore has an important interest in deciding which local feature descriptors have the most discrimination and the best matching performance. For over a decade, SIFT was possibly the most common feature descriptor for such activities.

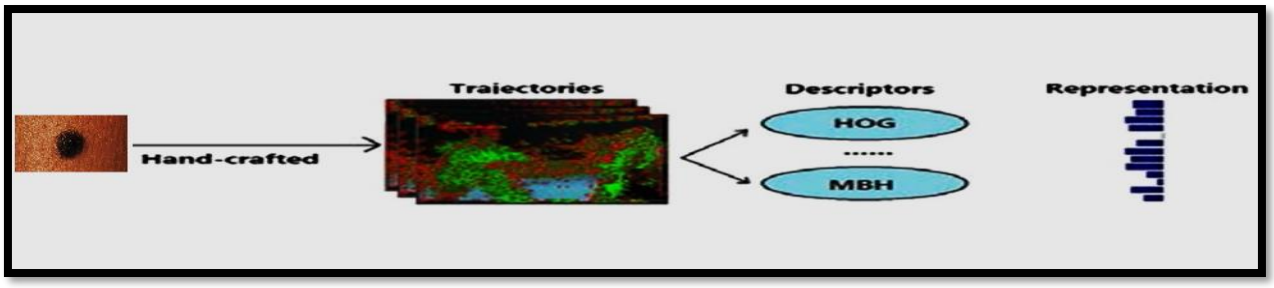

Figure 20: Work Flow of Hand-Crafted Feature Classifier [31]

5.1. Local Binary Patterns (LBP)

Local Binary Patterns (LBPs) are a very efficient approach to the extraction of features. It encodes distinctions between pixel intensities in the local pixel neighborhood. It is used in a variety of computer vision applications. For the recognition of the face, histograms of the LBP values are used. The face is usually divided into rectangular regions and the histograms are computed in each of them. The concatenation of the these histogram values is used as a representation of the face.

LBP is suggested successful descriptor of texture, which is characteristically local neighborhood texture, was the Local Binary Pattern. LBP encodes image pixels by thresholding the central pixel by using tiny 3 x 3 pixel image regions. If the center pixel has a gc value and the neighborhood pixels a gp strength (p = 0, 1, 7), the pixels are then measured [32].

The proposed LBP by Ojala, mainly due to its small computing complexities and the ability to code fine details, has soon become a common descriptor. The canonical operator LBP shall be calculated at each pixel location, taking into account the values of a small circular neighborhood with a R pixel radius, in the central pixel Ic.

$$\text{LBP}\ (N, R) = \sum_{n=0}^{N-1} s(I_n - I_c)2^n \quad [32]$$

When N is the number of pixels in the neighborhood, R is the radius, and s(x) = 1 if x 0, s(x) = 0 otherwise. The descriptor is the histogram of such binary numbers.

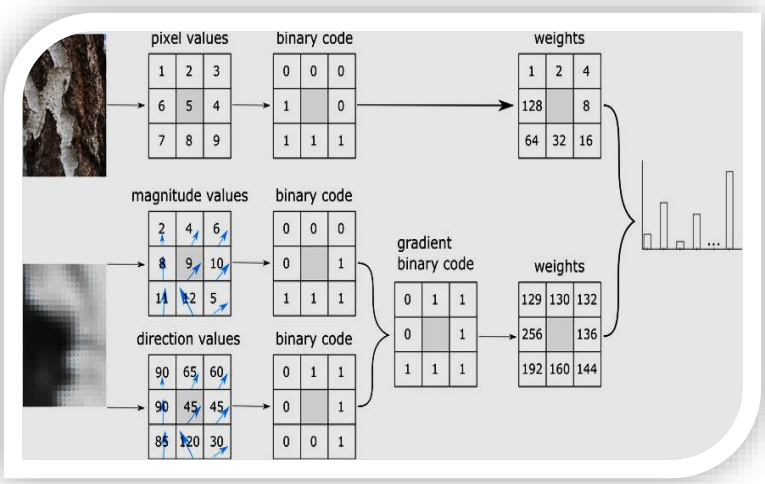

Figure 21: Local Binary Patterns applied on grey scale Images [32]

The image descriptor is displayed as the occurrence frequency (i.e., histogram) for all patterns collected in the image. When the LBP is computed for all pixels, a histogram describes its distribution over the whole image, given the space resolution of W / H pixels:

$$H = \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} f(LBP(i,j),p), p \in [0,N], \text{ [32]}$$

And

$$f(x,y) = \begin{cases} 1, \text{if} x = y \\ 0, \text{if} x \neq y, \end{cases} \text{[32]}$$

Where N shows the LBP pattern's maximum value. A total of 28=256 different patterns are represented by integer values in [0,255] for square 3 to 3 image areas with 8 neighbors. The final histogram is a vector of 256 dimensions that can be used to describe the images. The ability to capture fine grained image data is a benefit of the original LBP operator. The limitation of the descriptor to capture image details at different scales however is limited to a 3 to 3 dimension, In addition to vibration and noise response.

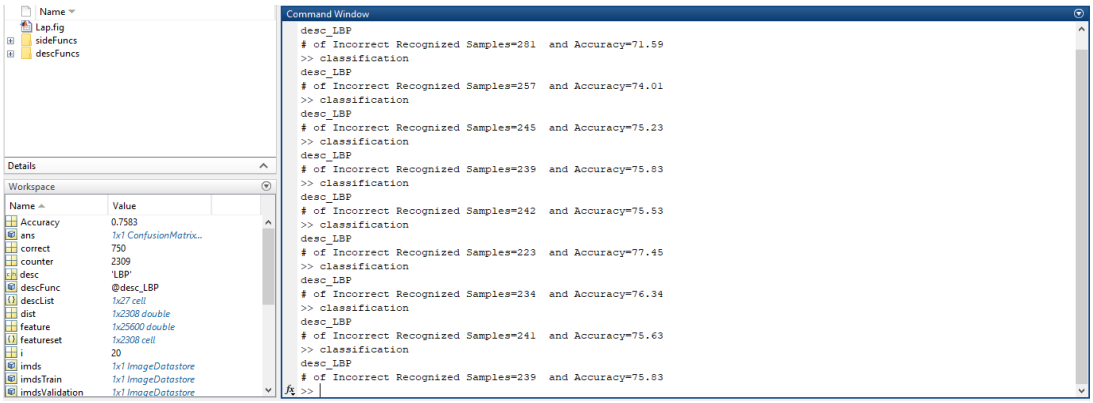

Figure 22: Results of LBP algorithm by using the ISIC dataset

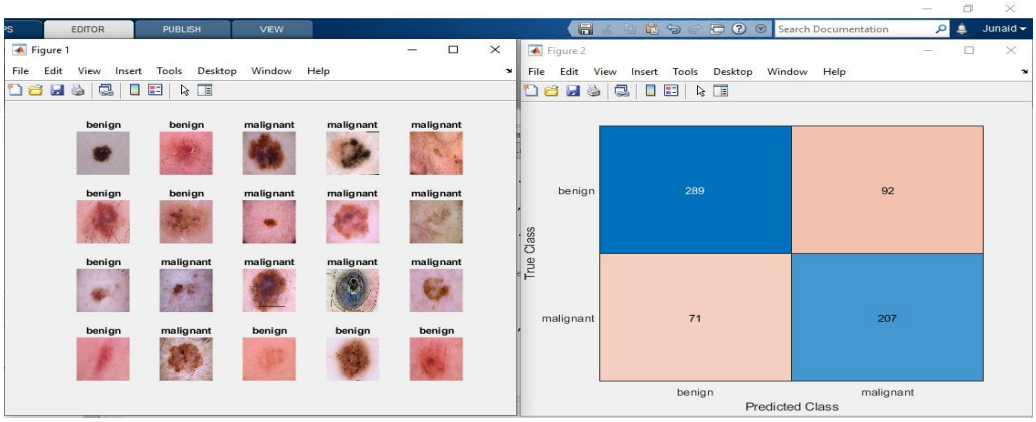

Figure 23: LBP confusion matrix

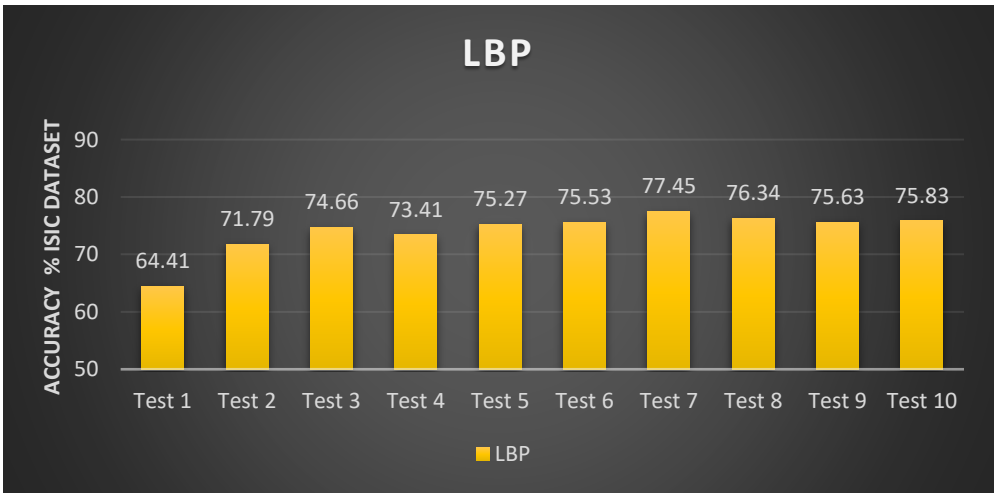

Chart 1: Accuracy of LBP algorithm using ISIC database with different values of Training and Testing data and Gridhist values.

5.2. Monogenic Binary Pattern (MBP)

MBP consists of two parts: one is a uniformly encoded magnitude, and another is a quadrant-bit-encoded monogenic orientation [33]. It is also one of the major used facial recognition feature descriptors. The magnitude is the measurement of local structural energy in the monogenic representation of an image and the conventional LBP operating device is used to encode local energy variation. The LBP operator can encode an 8-bit binary code with a local 3-3 patch. [33] The uniform LBP operator was subsequently suggested, which includes two bitwise transitions at most from 0 to 1 or reverse when the binary chain is regarded as circular. The LBP uses 6 bits to describe information on the local structure without significantly affecting its performance.

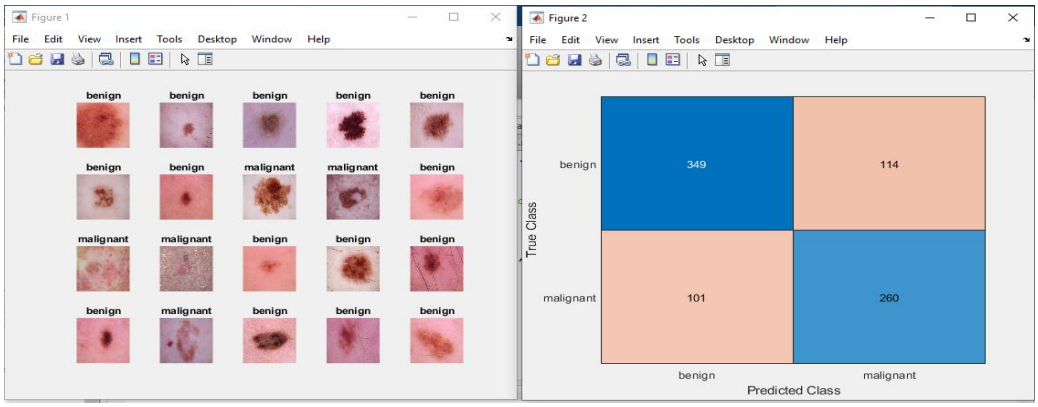

Figure 24: MBP confusion matrix

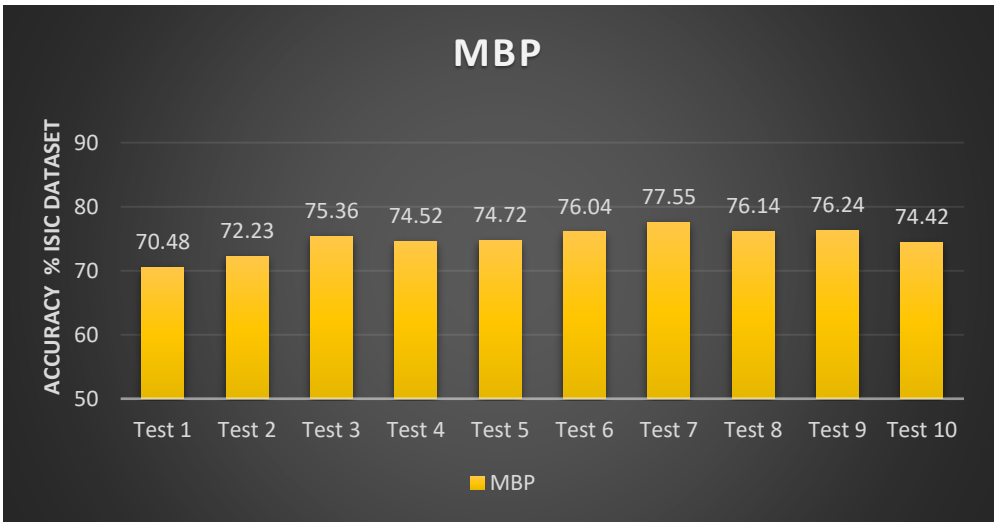

Chart 2: Accuracy of MBP algorithm using ISIC database

5.3. Local Gabor Binary Pattern Histogram Sequence (LGBPHS)

A face representation approach focused on non-statistics in which the training procedure for constructing the face model is unnecessary. In the LGBPHS-approach a face image is modeled as a "histogram sequence", [34] using multiple Gabor mapper and multi-oriented Gabor-filter. Each GMP is transformed into a Local Gabor binary (LGBP) map, and each LGBP map is further divided into non-overlapping re. In LGBPHS-approach, the front picture is a "histogram sequence" and Fig. 10 describes how features are extracted from greyscale images LGBPHS feature descriptor:

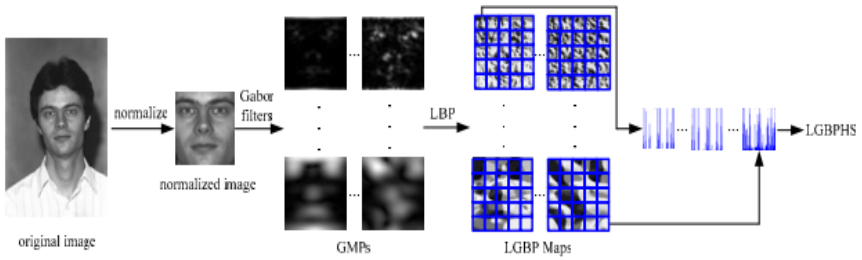

Figure 25: The framework of the proposed LGBPHS face representation approach [34]

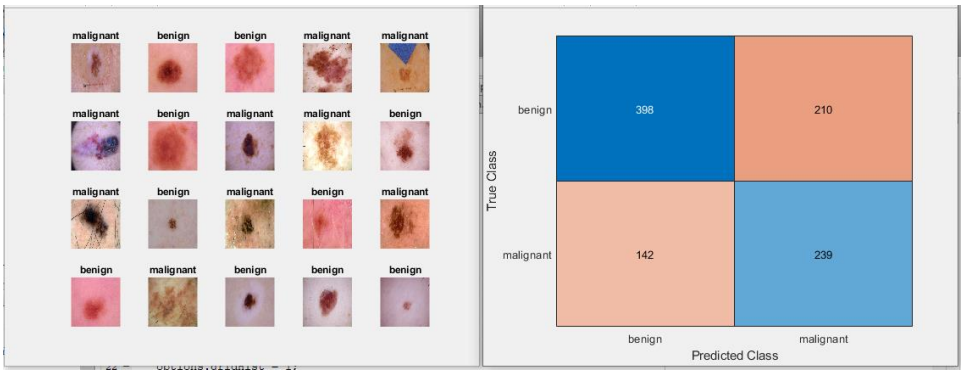

Figure 26: LGBPHS confusion matrix

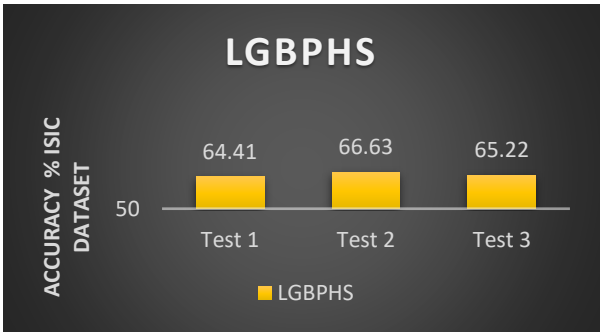

Chart 3: Accuracy of LGBPHS algorithm using ISIC database

5.4. Local Gradient Increasing Patterns (LGIP)

It is also one of the most famous for being used as facial recognition feature descriptor. Using eight binary bits, an LGIP feature encodes the intensity-growing trends in eight directions at each pixel, and then a decimal code is assigned to describe the overall trend that is increasing.

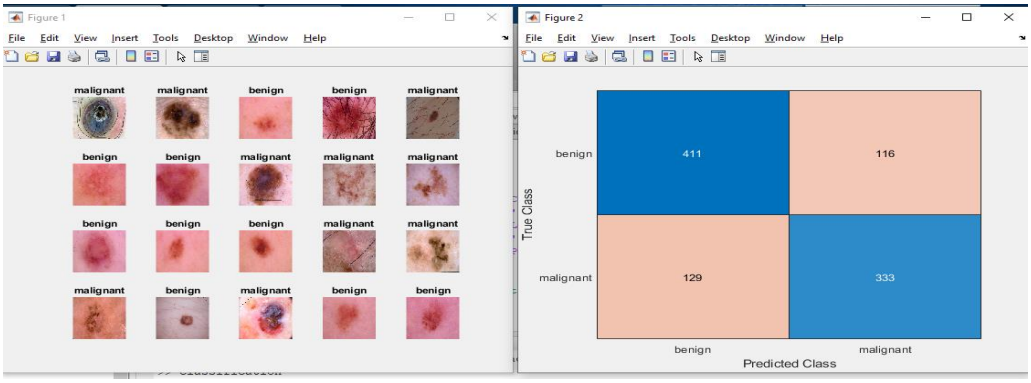

Figure 27: LGIP confusion matrix

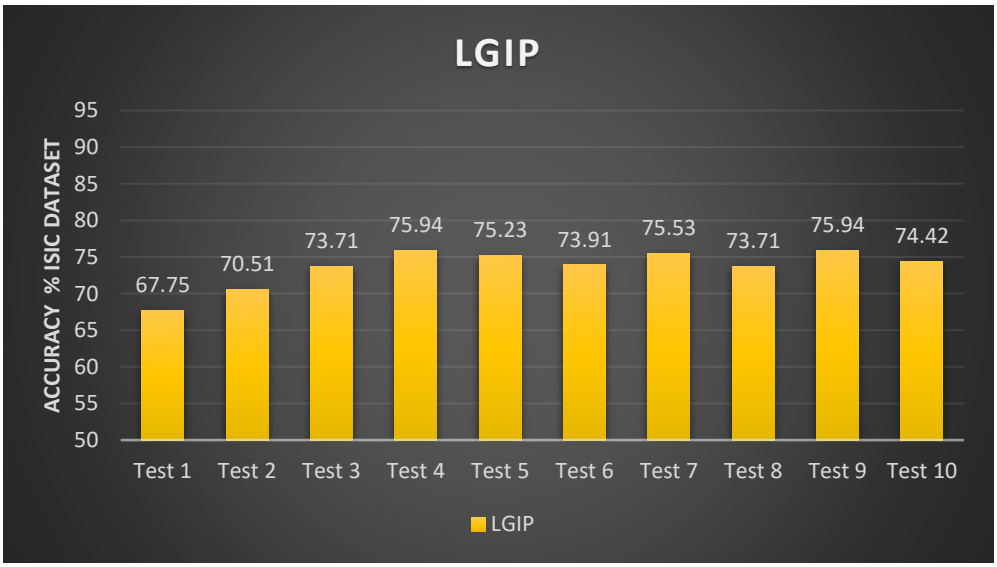

Chart 4: Accuracy of LGIP algorithm using ISIC dataset

5.5. Local Gradient Patters (LGP)

It is also one of the most famous for being used as facial recognition feature descriptor. Local pattern is calculated by the local gradient flow through the central pixel in the region of 3x3 pixels from the one side to the other. Two separate two-bit binary patterns, known as the LGP for the pixel, represent the center pixel of the region.

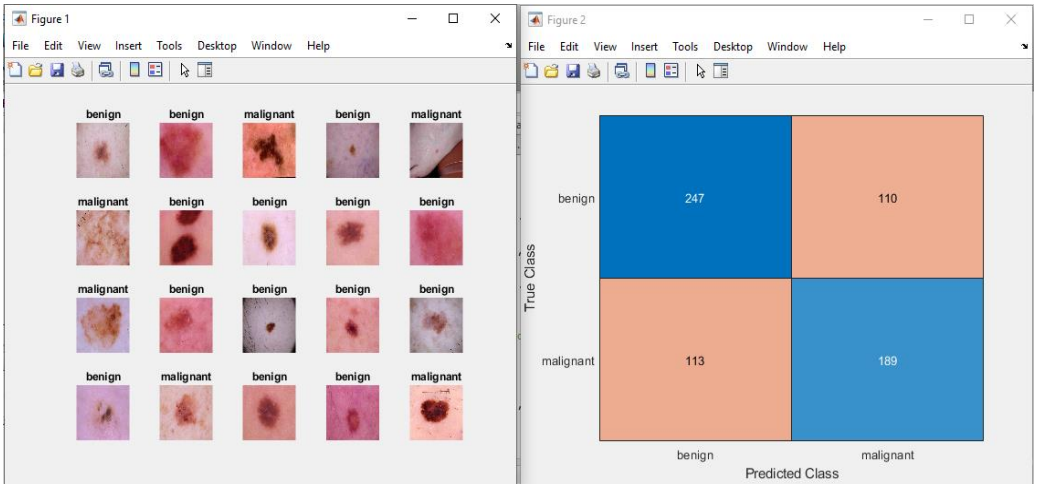

Figure 28: LGP confusion matrix

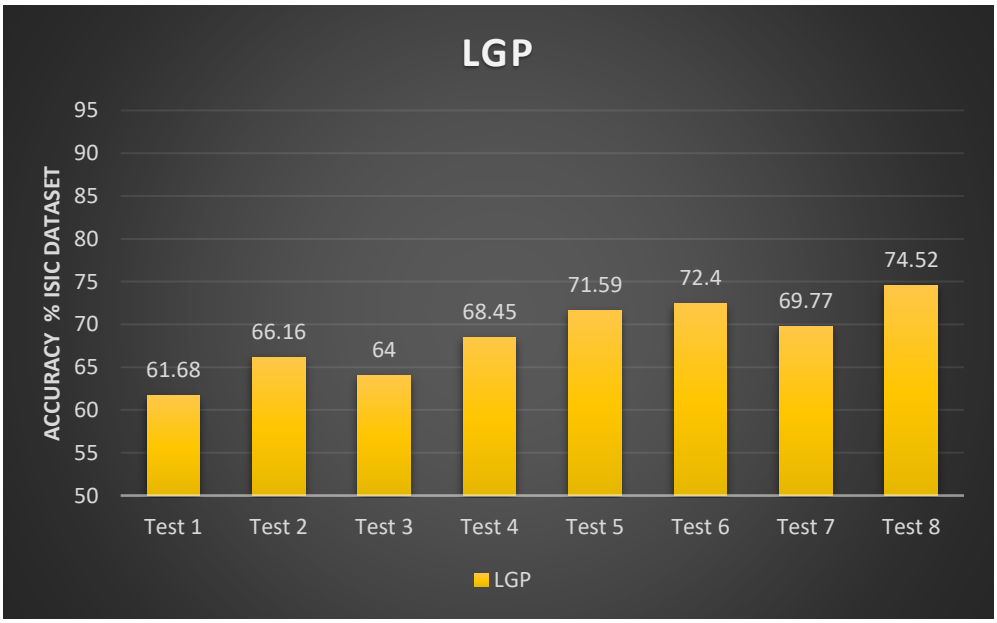

Chart 5: Accuracy of LGP algorithm using ISIC dataset

5.6. Monogenic Binary Coding (MBC)

In three complementary components, monogenic signal representation decomposes the original signal: amplitude, direction, and phase [35]. In each local region monogenic variation and in each pixel, a monogenic variation is encoded and then the statistic features (e.g., histogram) of the local characteristics extracted are calculated. The local statistical features extracted from the complementary monogenic components are then fused into an effective FR (amplitude, orientation, and phase) [35].

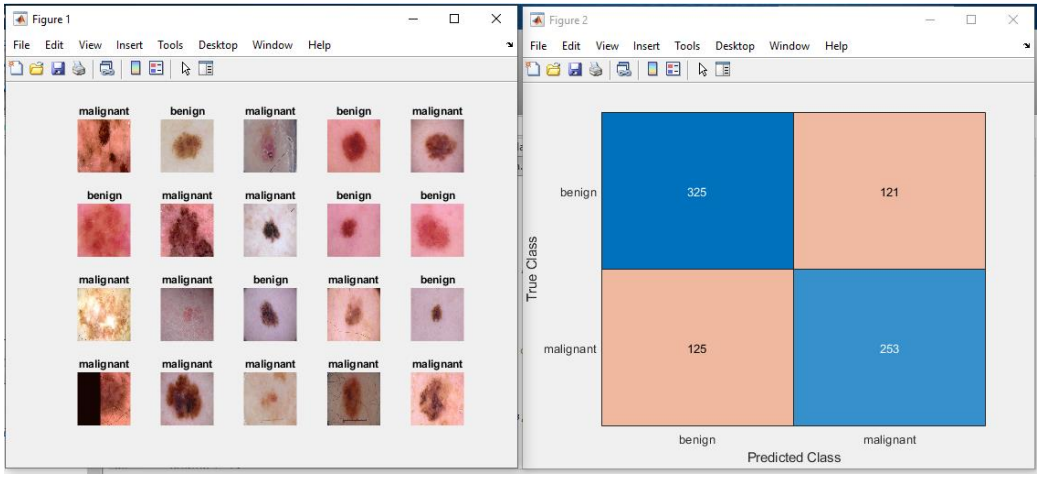

Figure 29: MBC confusion matrix

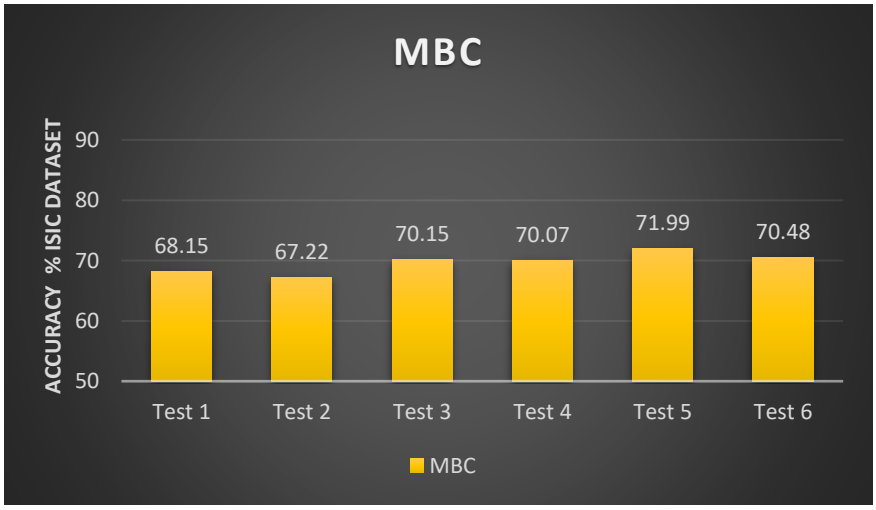

Chart 6: Accuracy of MBC algorithm using ISIC dataset

5.7. Median Robust Extended Local Binary Patterns (MRELBP)

MRELBP compares local image medians instead of brute image intensities in contrast to traditional LBP and many LBP variants. Median Robust Extended Local Binary Pattern are considered as one of the finest texture classification feature descriptors. The local binary model suggested by Ojala ET AL. [36] Is characterized by encoding the differences between the central pixel value and the neighboring points, only taking into account the sign information in order to form a local binary pattern.

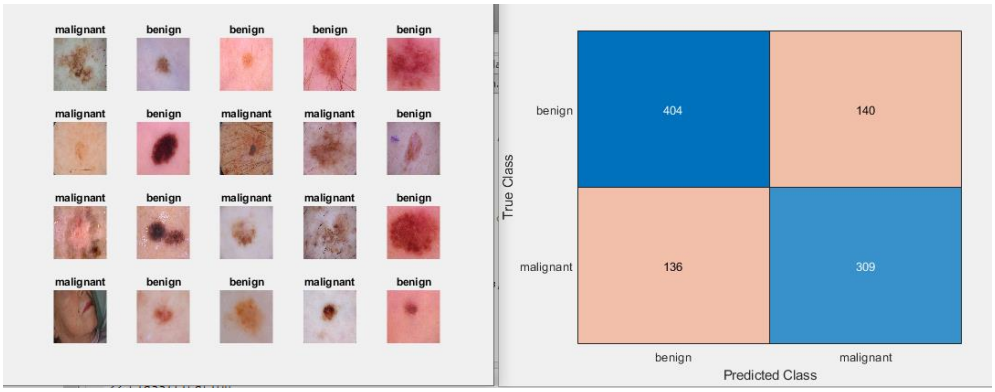

Figure 30: MRELBP confusion matrix

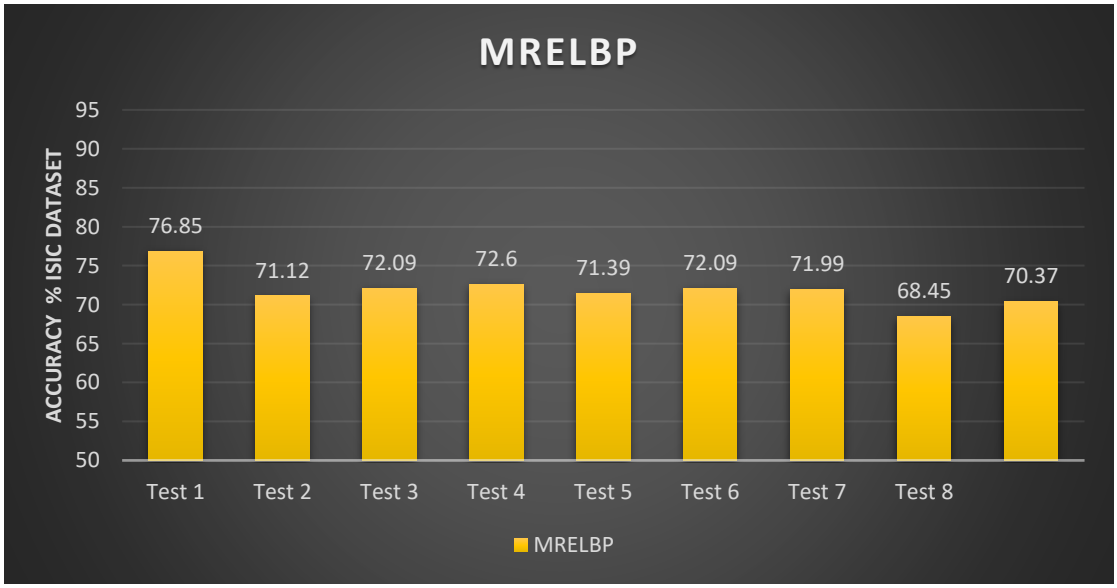

Chart 7: Accuracy of MRELBP algorithm using ISIC dataset

5.8. Median Ternary Pattern (MTP)

MTP encodes local neighborhood texture information by quantizing the neighbor's intensity values in three different rates around each pixel based on the local median and global threshold value. A facial feature descriptor is to distribute the resulting MTP codes within an image or face patch. In the presence of noise, median filters are efficient. Therefore, a system that uses the median is stronger than noise and therefore produces reliable information. Due to noise or light differences the intensity of a pixel can vary [37]. Deploy a window (threshold ±t) to construct a 3-value code called a ternary pattern instead of using a 2-value number. This reduces the noise effect in nearly uniform regions. Next, the medium strength of the nine pixels is measured for each pixel around a neighbor 3 X 3 [37][38]. Then the median M within ±t is set to 0, then MTP code comes out as:

$$S_{MTP}(v) = \begin{cases} 1, v > M_c + t \\ 0, M_c - t \leq v \leq M_c + t \\ -1, v < M_c - t \end{cases} \text{ [38]}$$

Image localization using MTP feature descriptor is shown in Fig. 19 that how features are first divided into patches and then illustrated in localized histograms.

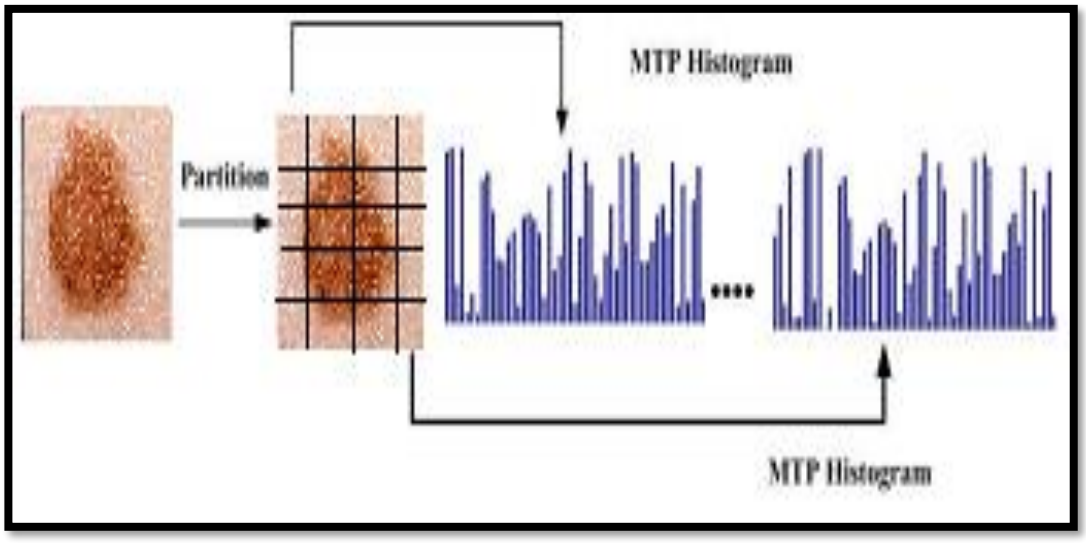

Figure 31: Image localization using MTP [38]

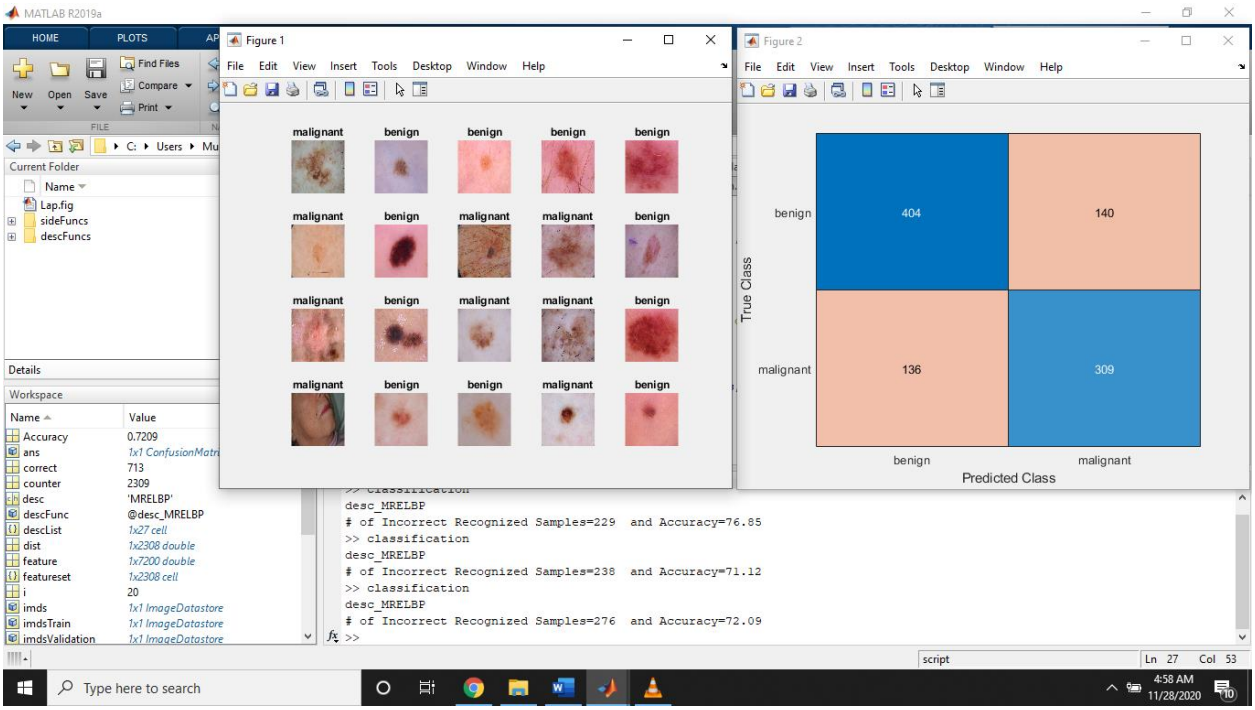

Figure 32: MTP confusion matrix

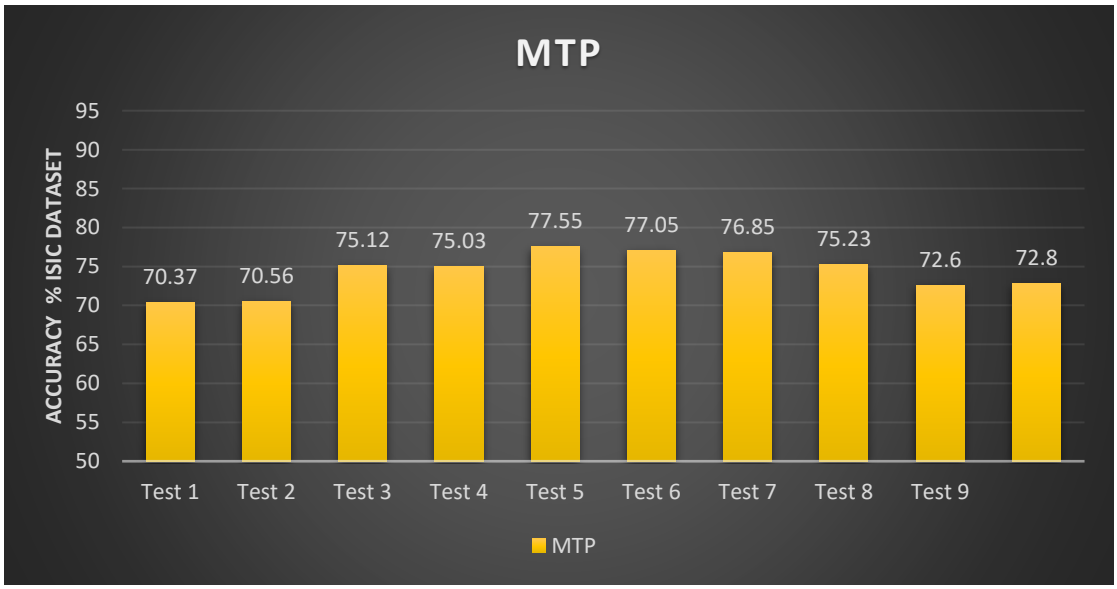

Chart 8: Accuracy of MTP algorithm using ISIC dataset

5.9. Weber Local Descriptor (mWLD)

Weber Local Descriptor (WLDs), is focused not just on the shift in the stimulus (such as sounds, light) but on the original strength of the stimulus, as a consequence of humans perceiving the pattern. WLD is comprised of two components: differential excitation and orientation. The differential excitation aspect is the function of the relationship between two terms: one is the relative intensity of a current pixel and its neighbors; the other is the current pixel intensity. The orientation part is the current pixel's gradient orientation [39]. For one particular picture, we create a WLD histogram using the two components.

The two WLD differential excitation / orientation components (den) are defined. Following this, we will present how to compute an image (or image region) for a WLD histogram. We use the intensity differences between its neighbors and a current pixel to adjust the current pixel in Differential Excitation. Thus, we hope to find the essential variants in a picture that simulate the perception of the pattern of human beings. In particular, as illustrated, the differential excitation (xc) of the current xc pixel is determined. The variations between its neighbors and the center point are first determined using the filter. [39] Drop down architecture of Mwld feature descriptor including all parameters like filtering, labeling etc. is illustrated in Fig. 21:

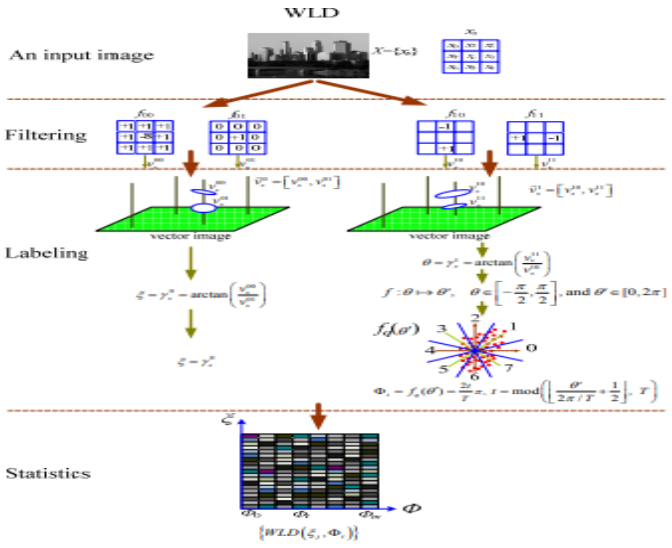

Figure 33: Drop Down WLD Feature extractor Architecture [39]

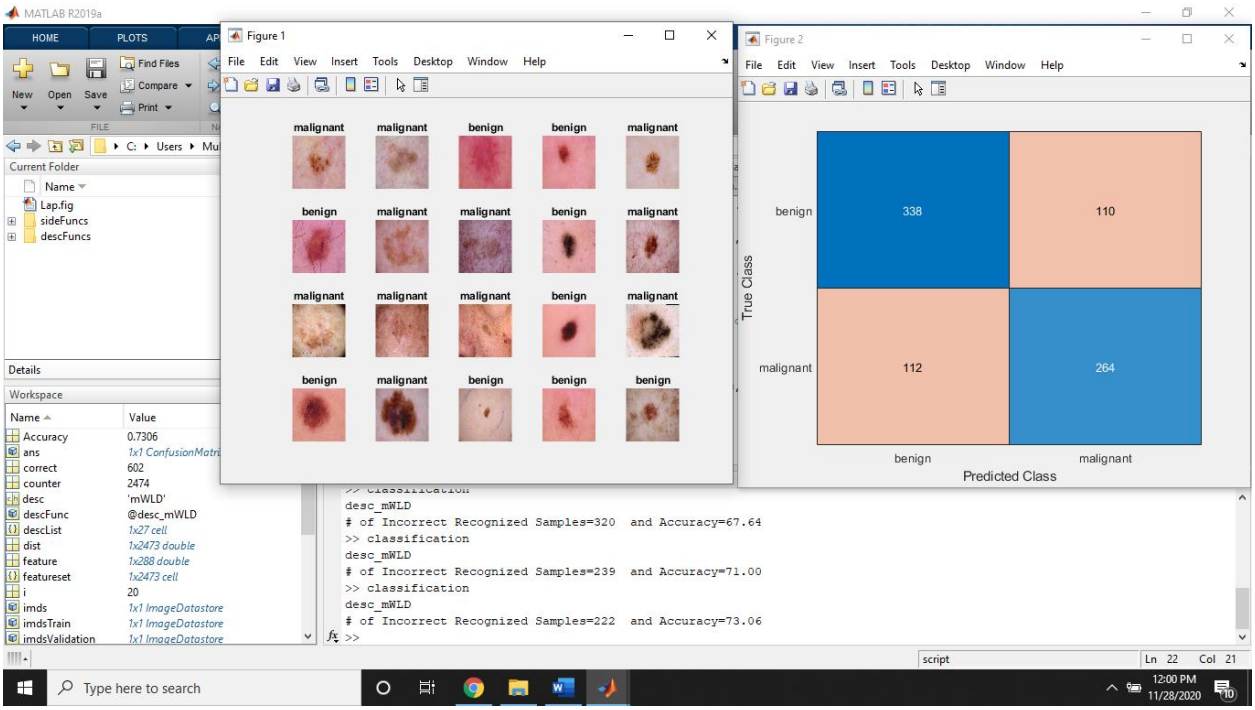

Figure 34: mWLD confusion matrix

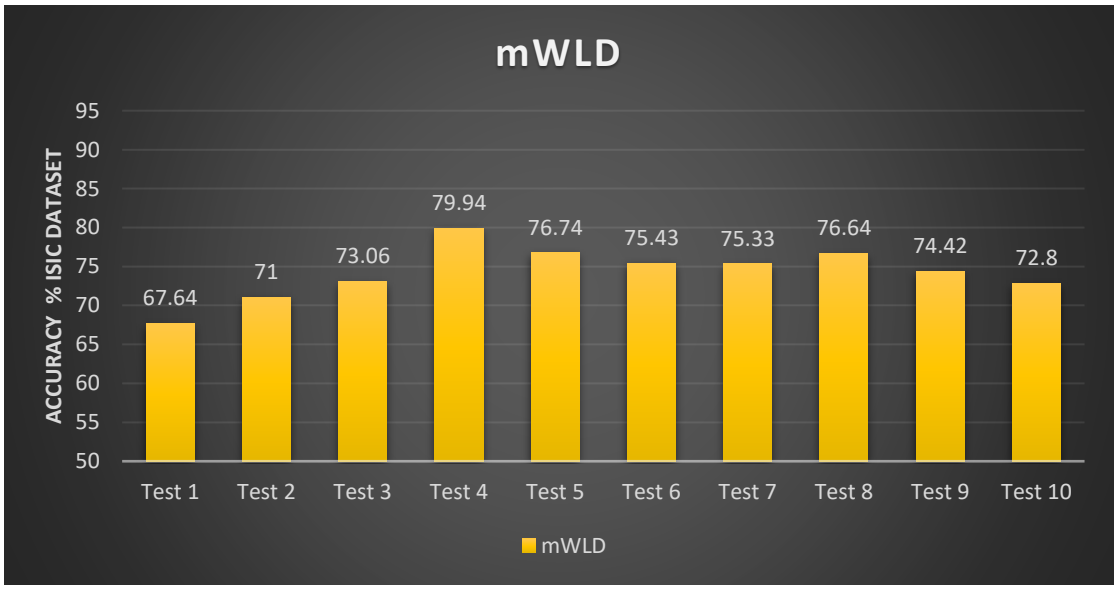

Chart 9: Accuracy of mWLD algorithm using ISIC dataset

5.10. Local Directional Number Pattern (LDN)

LDN is a facial descriptor that evaluates local neighborhood directional information for encoding its structure. LDN is primarily a texture descriptor and encodes structural details facial structure and variations in intensity [40]. The first step for the formation of LDN is generation of Gaussian edge responses and they are accomplished through the Gaussian mask. These Gaussian masks are placed over the images of the face, and are combined to create the images of the edge response. The LDN histogram consists of the coded LDN image and contains different features of potential points, edges, curves, and other facial characteristics of a local texture. The histogram does not encode location information, but encodes event information. The face image is therefore divided into small sections for each section and the histogram is extorted [40]. All histograms are combined, forming a histogram of the LDN.

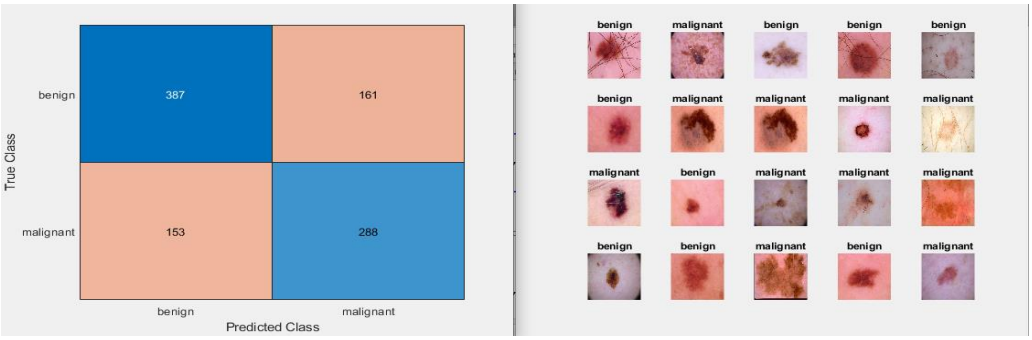

Figure 35: LDN confusion matrix

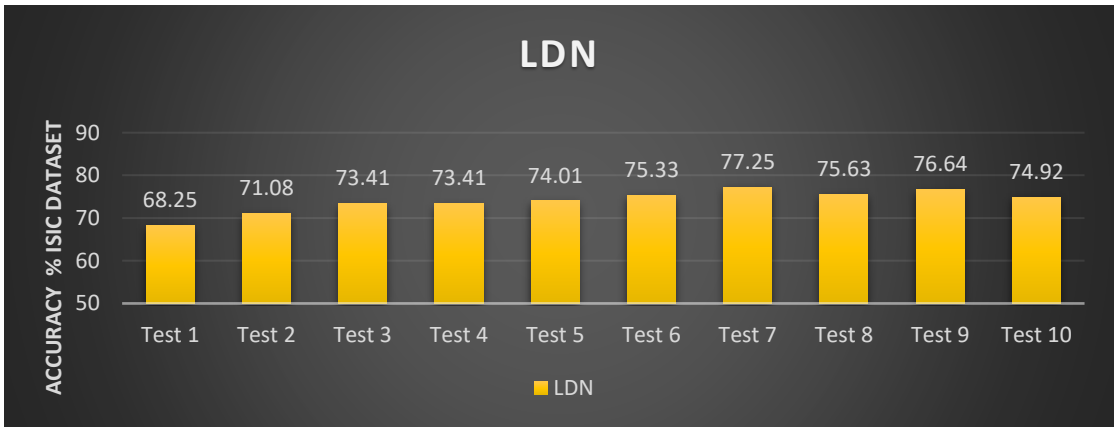

Chart 10: Accuracy of LDN algorithm using ISIC dataset

5.11. Binary Pattern of Phase Congruence (BPPC)

The suggested descriptor is a multi-scaling local descriptor known as the Binary Pattern of Phase Congruence (BPPC) that can represent a variety of facial image patterns. It is rendered on directed PC images by applying LBP.

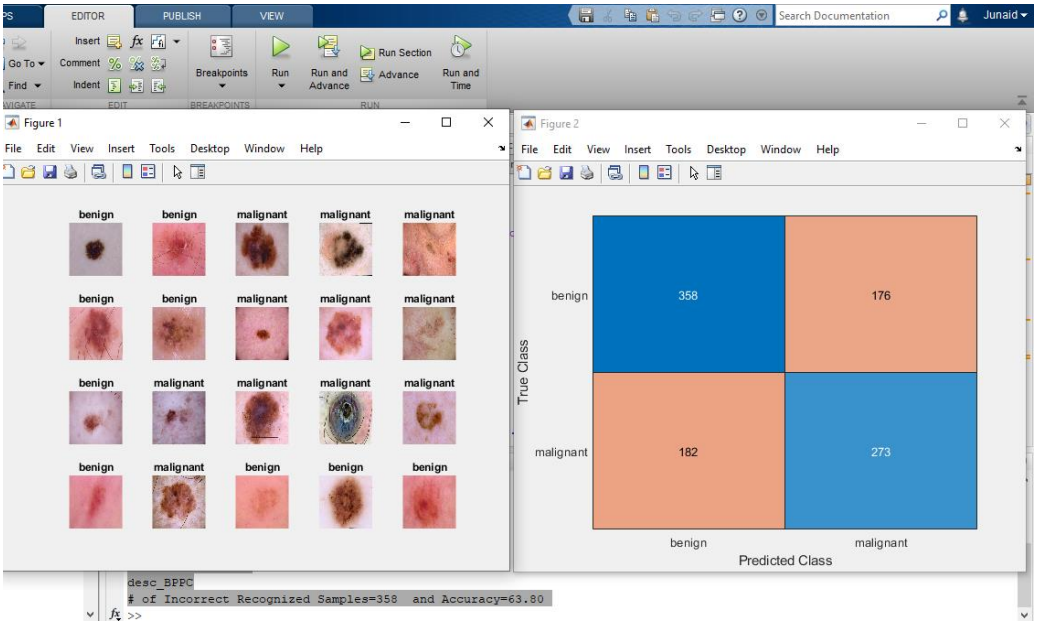

Figure 36: BPPC confusion matrix

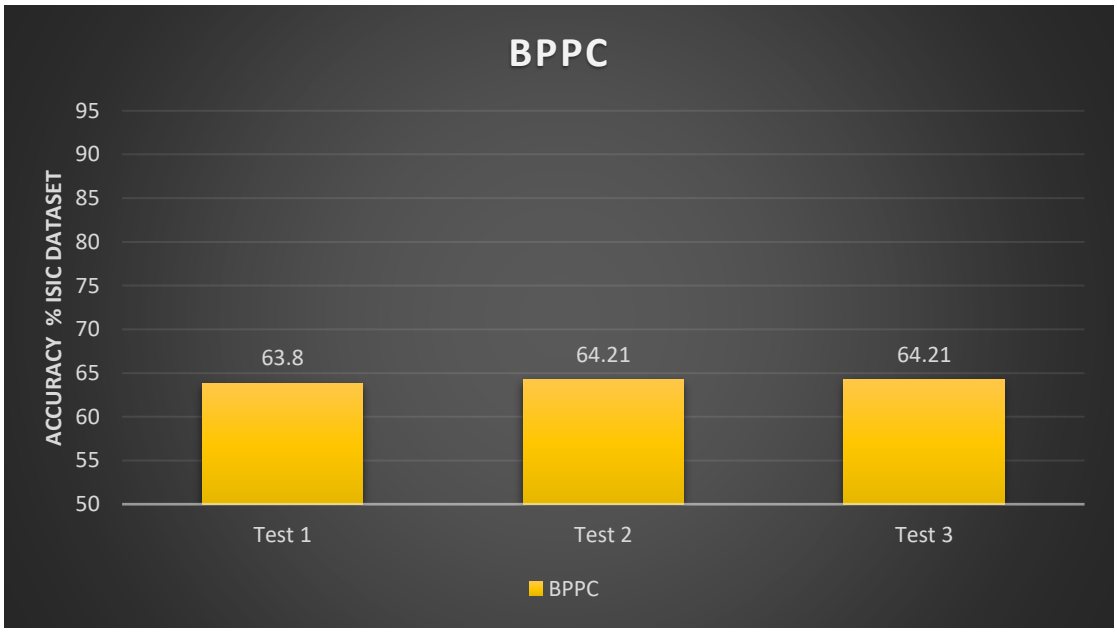

Chart 11: Accuracy of BPPC algorithm using ISIC dataset

5.12. Gradient-Based Patterns (GDP)

By quantizing the gradient directional angles, the GDP operator encodes the texture information of a local area to form a binary pattern. The GDP's micro-samples are then used as the feature descriptor for their location and appearance detail [41]. The resulting GDP features retain more information than the grey-level approaches and define local image elements in a more stable manner, as the gradient operator can effectively boost the edge information of an image.

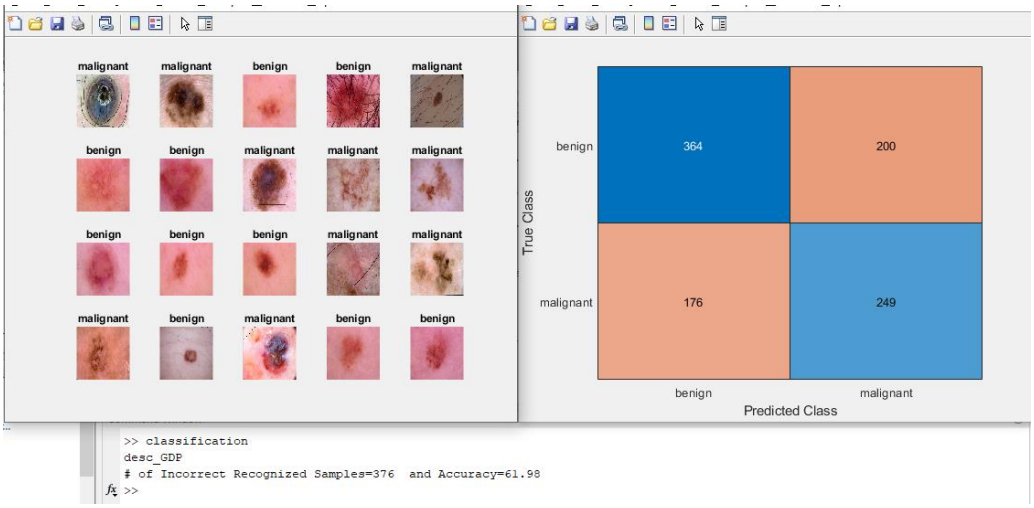

Figure 37: GDP confusion matrix

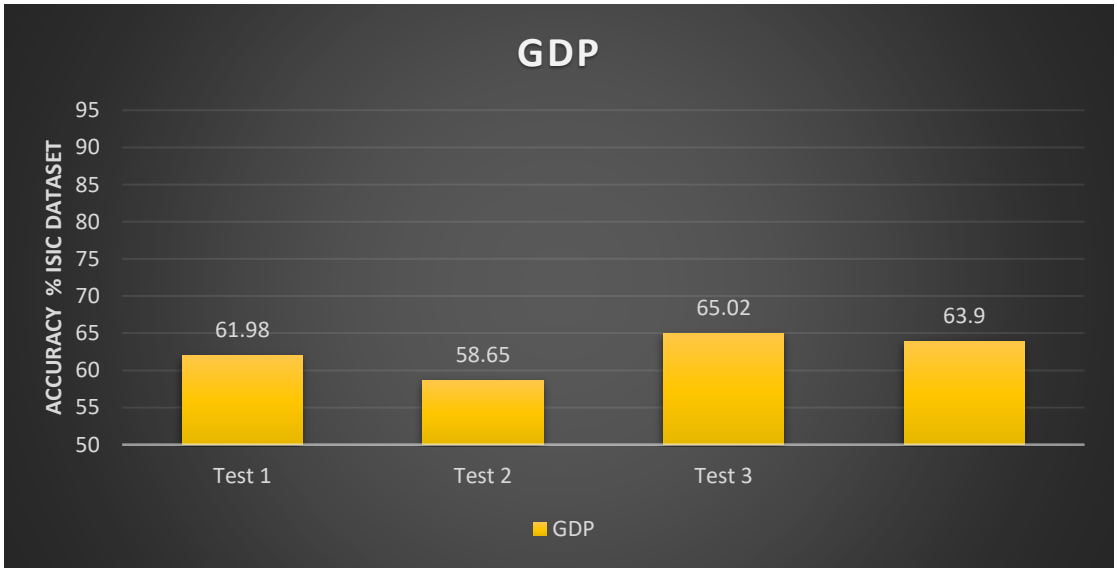

Chart 12: Accuracy of GDP algorithm using ISIC dataset

5.13. Local feature descriptor (LFD)

Local Feature Descriptors (LFD) are an increasingly efficient image comparison technique that is currently the best solution to in-plane rotation, distortion, and light changes [42]. However, it is important to remember those assumptions. It is also known as appearance feature. It is still more stable than a pixel feature, but the appearance property still disturbs the LFD ability. First, the feature is sensitive to light variations and the ability to explain the appearance depends on its complexity.

LFDs overcome the intrusion of the image variations by collecting features through the main areas. That can be seen as a kind of form description extension. However, LFD could not do anything for objects with low appearance complexity [42]. LFD does not collect any useful ball details for this scenario because the ball's differentiation is not the presence of the shape.

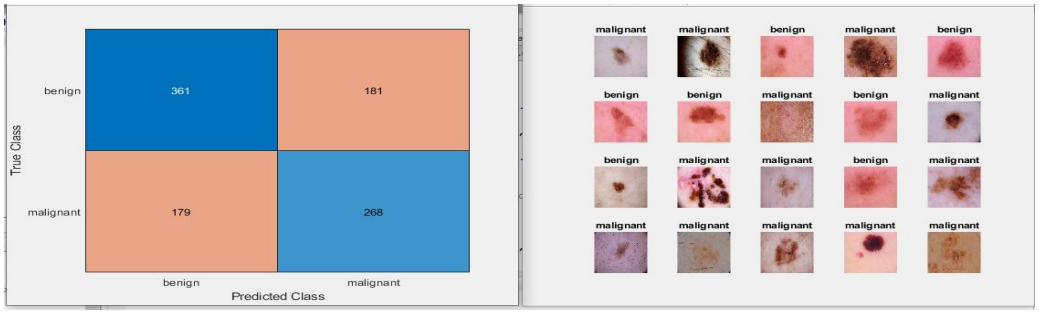

Figure 38: LDF confusion matrix

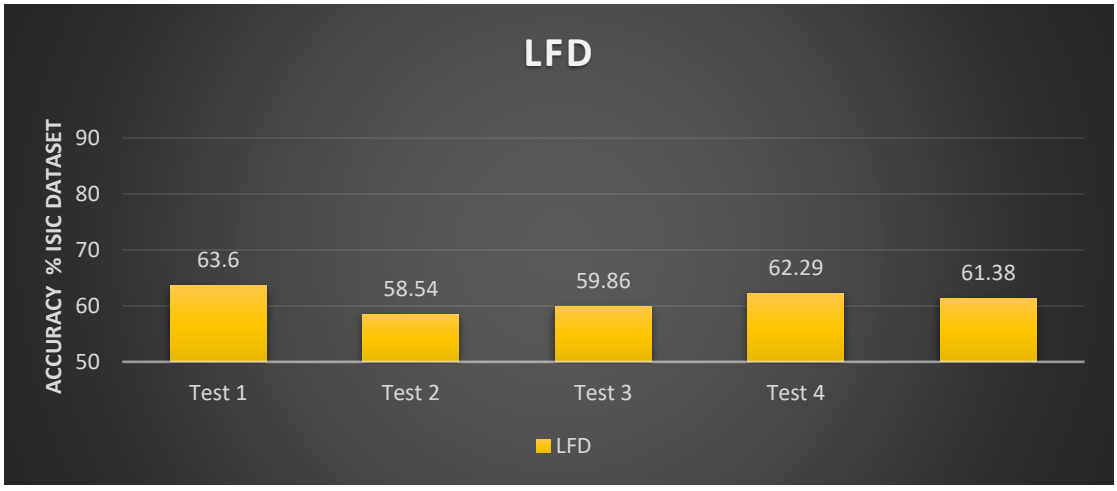

Chart 13: Accuracy of LDF algorithm using ISIC dataset

5.14. Local Directional Pattern (LDP)

A descriptor used for object detection is the Local Directional Pattern (LDP). For each pixel in the image, it assigns a code, and the resulting LDP image is split into regions for which a histogram is created [43]. The histogram bins are linked to the final descriptor of all regions. Slightly modification of LDP coding pattern constraints results in the redesign of patterns. DR-LDP's significance for effective object recognition is the compact code generation [43].

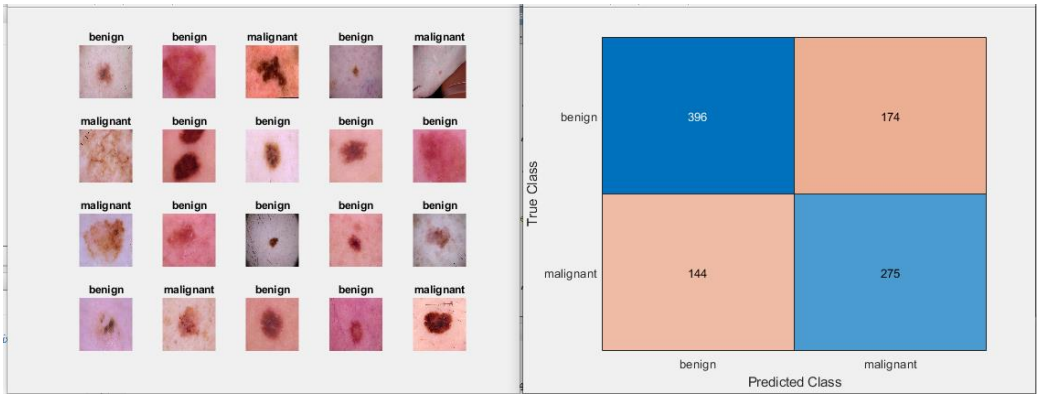

Figure 39: LDiP confusion matrix

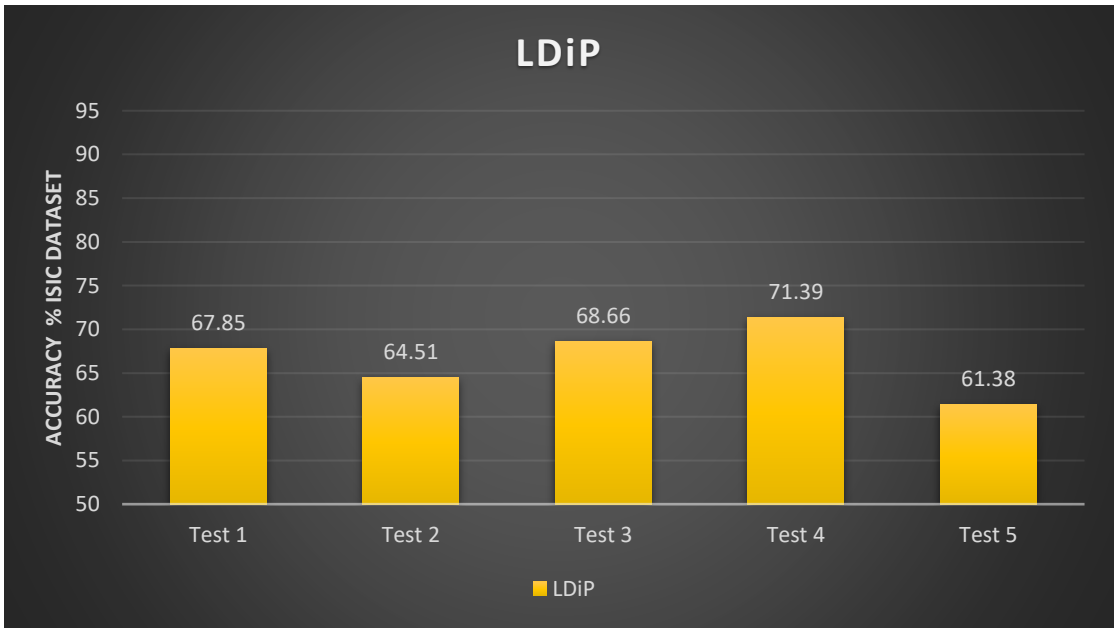

Chart 14: Accuracy of LDiP algorithm using ISIC dataset

5.15. Local Directional Ternary Pattern (LDTP)

The ad-hoc nature of the LDP code overlooks the effect of the neighborhood information since the code in the neighborhood contains all the information. In order to prevent these shortcomings, we are explored a new coding scheme that blends LDP's and LTP's 135 values. The local texture modeling proposed is based on local derivative variations and the encryption of contrast information, and inherits relevant characteristics [41,43]. The suggested descriptor known as LDTP integrates 2 separate types of compass masks: derivative-Gaussian masks to avoid distortion of the noise, make the device resilient for lighting changes.

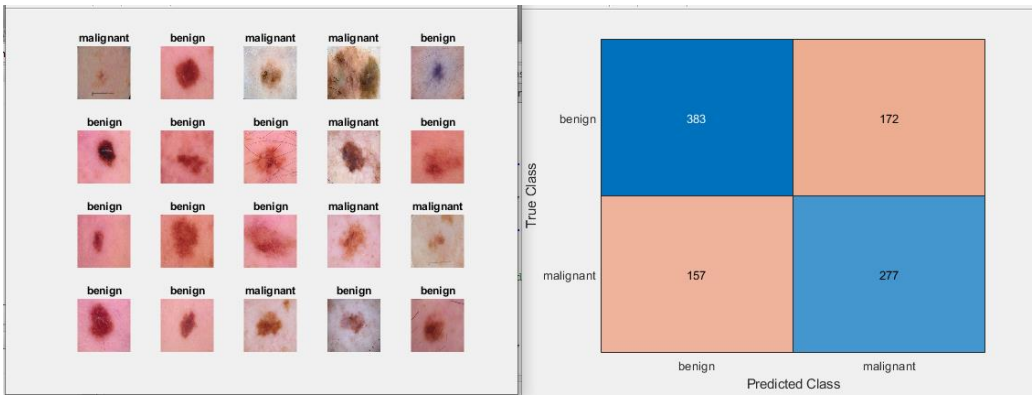

Figure 40: LDTP confusion matrix

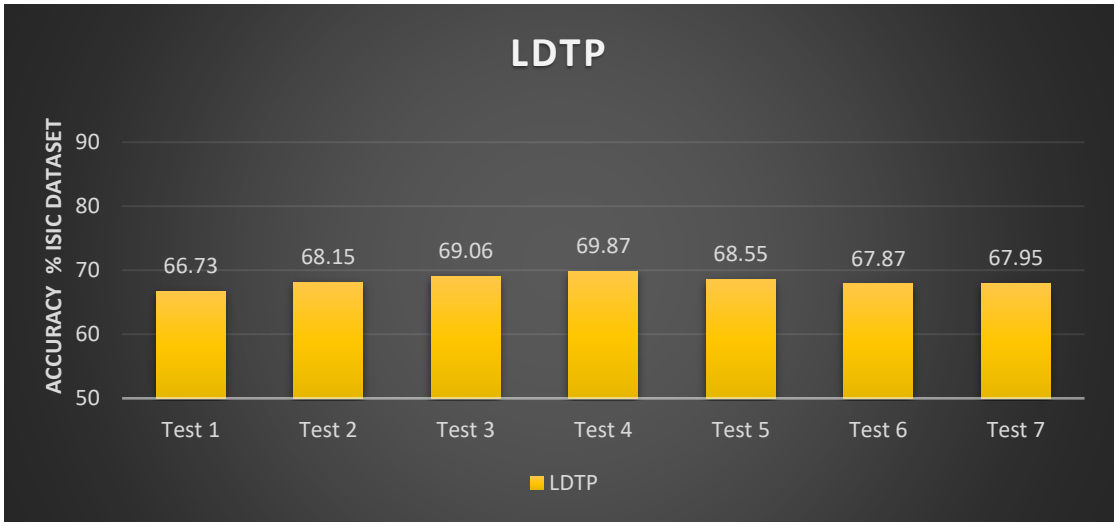

Chart 15: Accuracy of LDTP algorithm using ISIC dataset

5.16. Generalized Local Ternary Patterns (GLTP)

GLTP operator encodes local texture information using three separate types of discrimination, quantizing the gradient magnitude values of a particular neighborhood. The proposed encoding system can differentiate between smooth and highly textured sections and ensure that micro patterns in texture are created that are compatible with local image features (smooth or high-texture). The output of the GLTP function descriptor is evaluated empirically with a Support Vector Machine (SVM) [44].

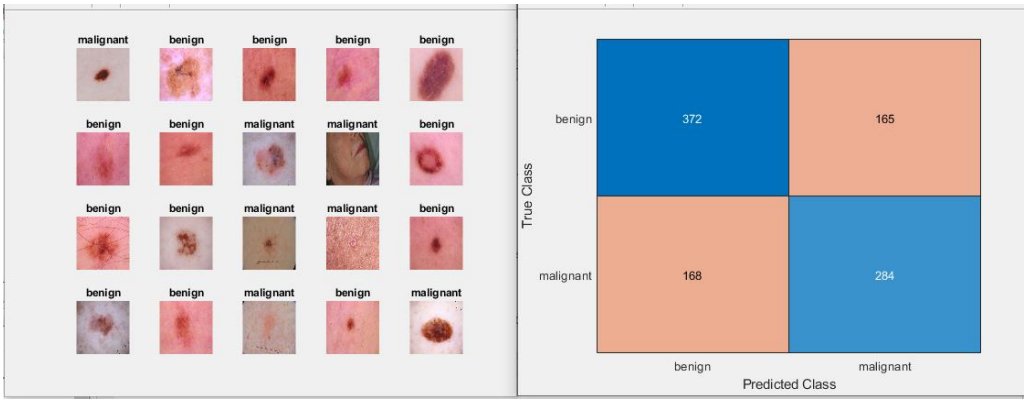

Figure 41: GLTP confusion matrix

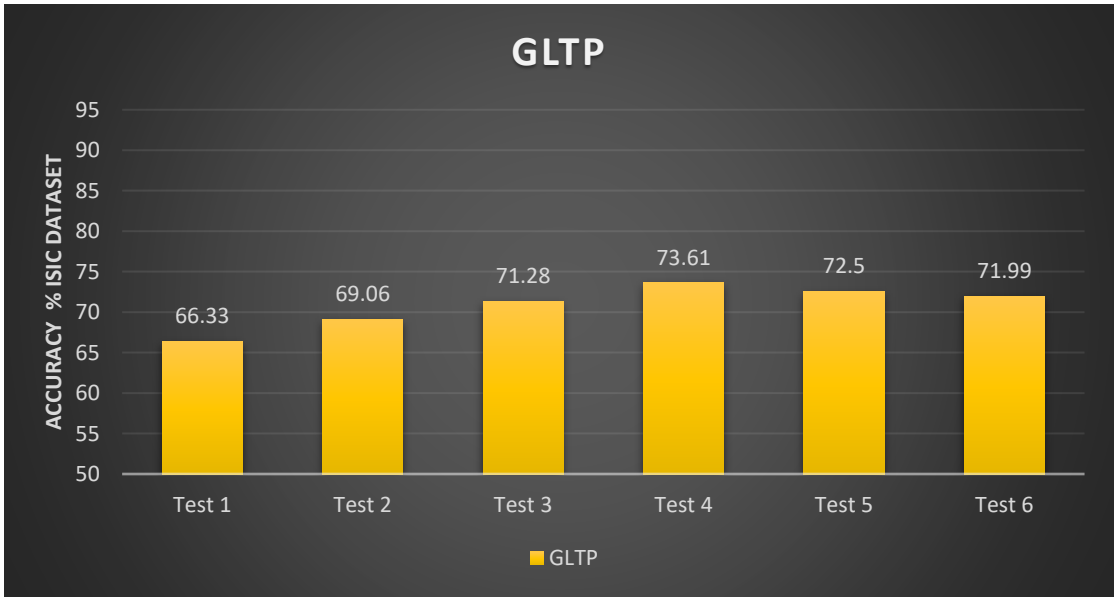

Chart 16: Accuracy of GLTP algorithm using ISIC dataset

5.17. Improved Weber Binary Code (IWBC)

IWBC is more polished form of IWBD and IWLD, combining both since they represent local patterns more efficiently and accurately by using weber magnitude and orientation components.

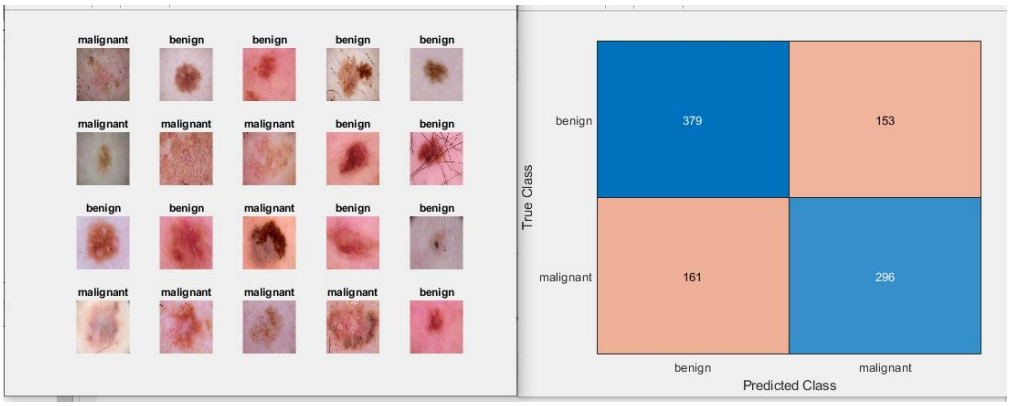

Figure 42: IWBC confusion matrix

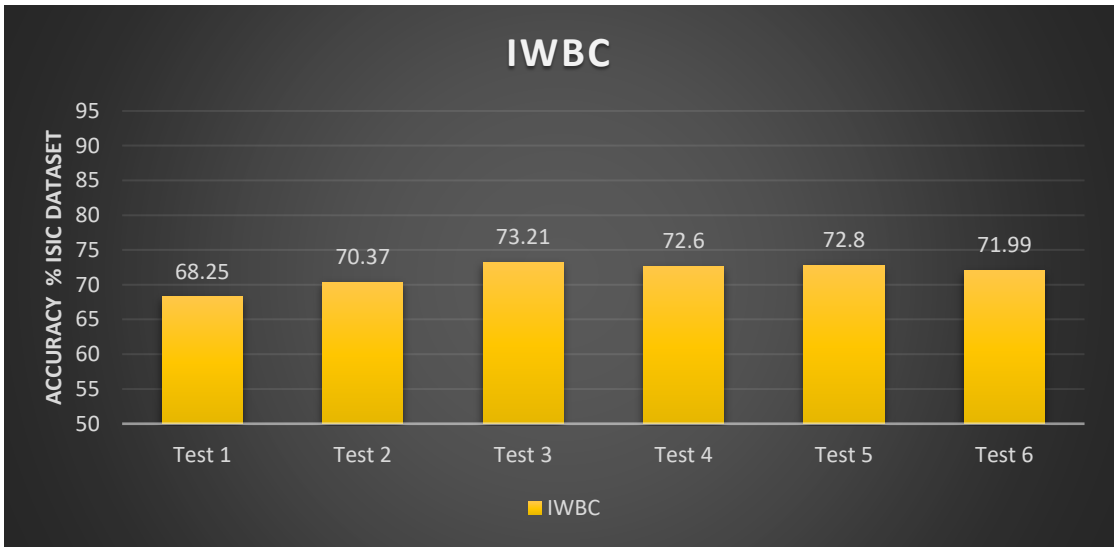

Chart 17: Accuracy of IWBC algorithm using ISIC dataset

5.18. Local Arc Pattern (LAP)

Local arc pattern is a feature based on grey color intensity values, dividing each image into equal sized blocks and histograms of LAP codes then quantizing them to build a classification feature.

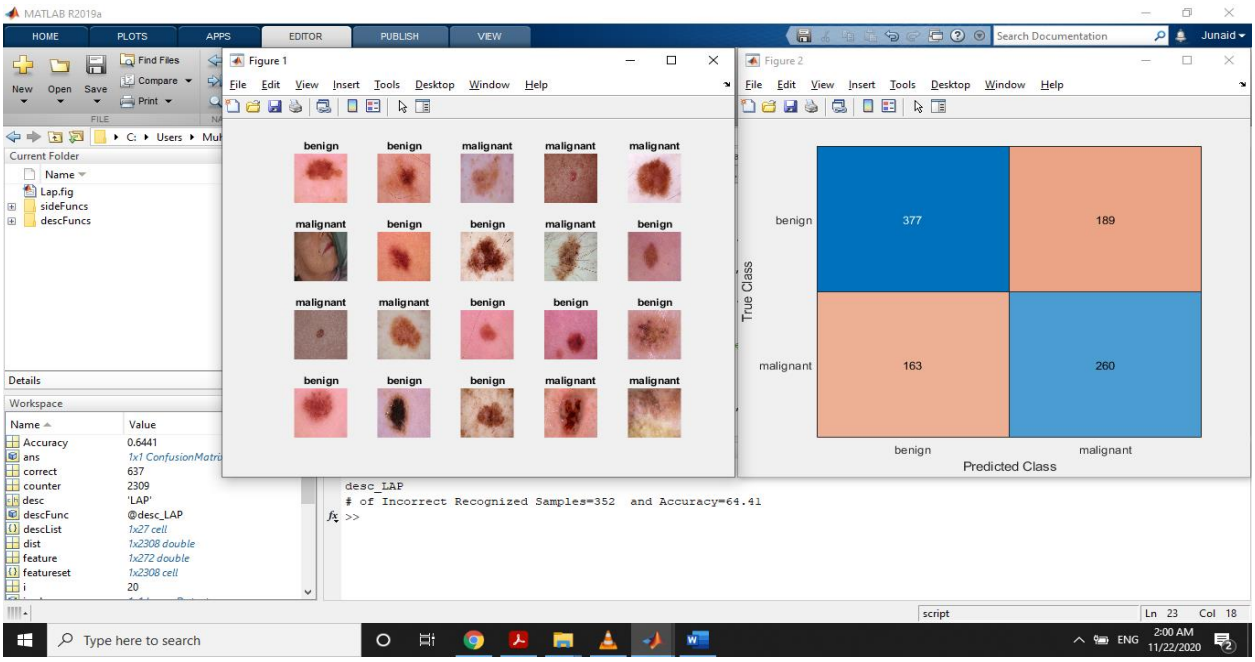

Figure 43: LAP confusion matrix

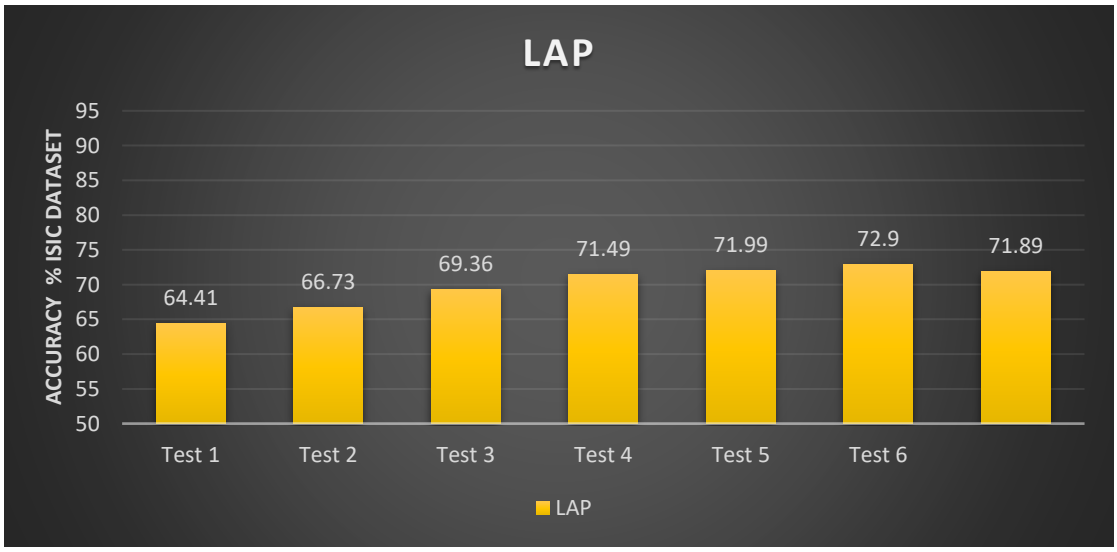

Chart 18: Accuracy of LAP algorithm using ISIC dataset

5.19. Local Direction Pattern Variance (LDiPV)

LDiPV is a more polished descriptor as compared to LDP, LDiPV introduces a local variance of directional response in order to encode the contrast data within the same descriptor.

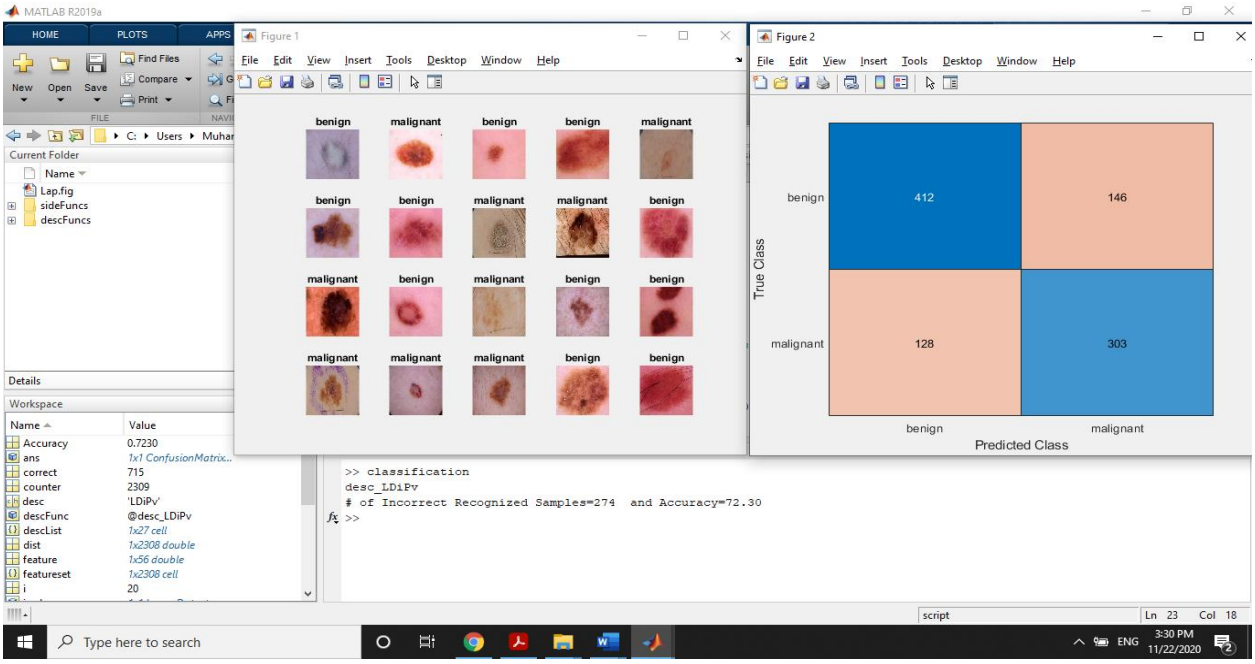

Figure 44: LDiPV confusion matrix

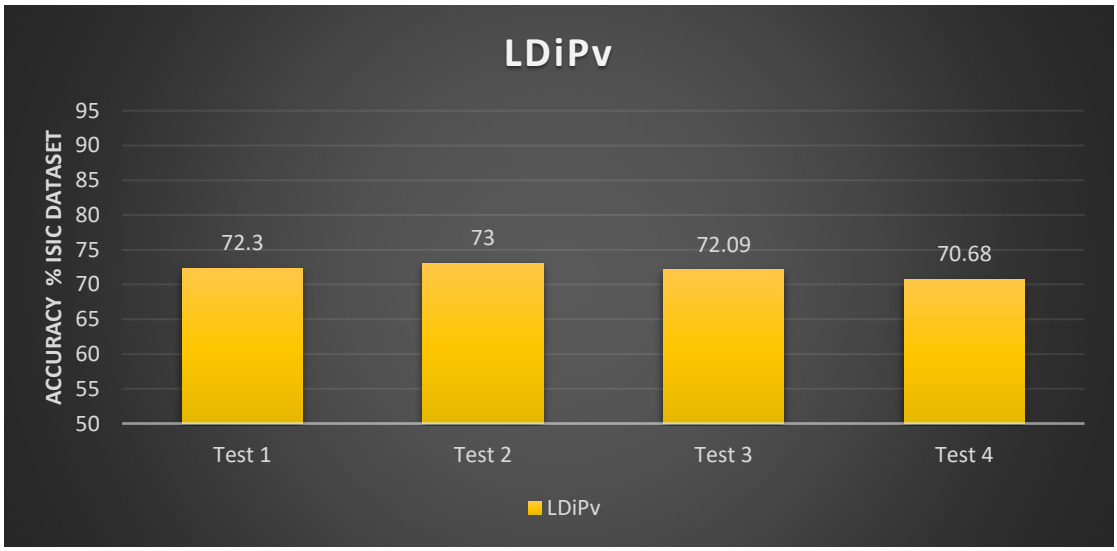

Chart 19: Accuracy of LDiPV algorithm using ISIC dataset

5.20. Local Monotonic Pattern (LMP)

Local Monotonic Pattern (LMP) can extract strong facial features from a face image that provides accurate and reliable expression recognition efficiency. The LMP operator applied to a pixel finds the monotonic intensity change of the neighboring pixel at different radii [45]. The micropatterns therefore find enhancement by carving the images of each tile and taking histograms with spatial information.

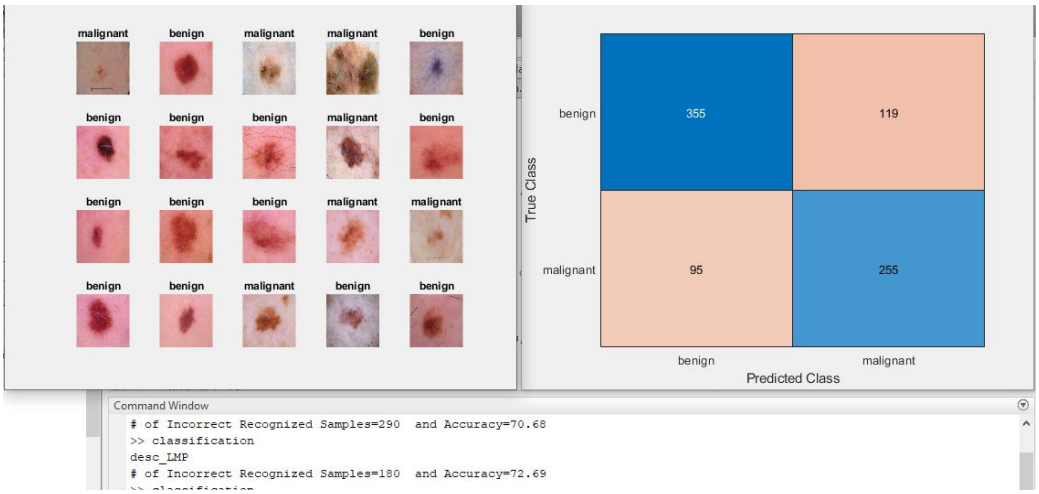

Figure 45: LMP confusion matrix

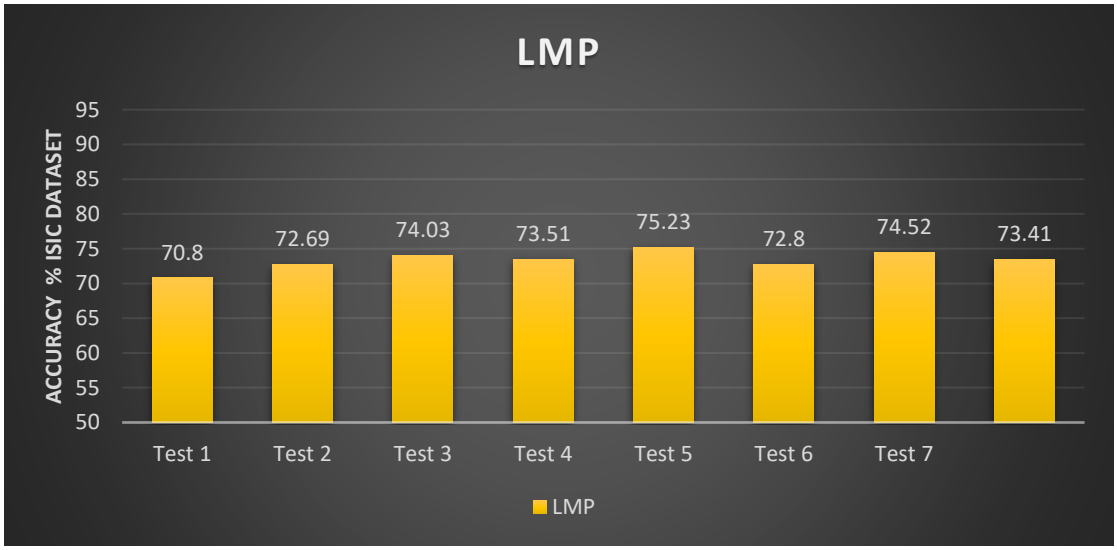

Chart 20: Accuracy of LMP algorithm using ISIC dataset

5.21. Local Phase Quantization (LPQ)

LPQ (Local Phase Quantization) is used as an image descriptor [46]. LPQ relies on a local image frame to compute the short-term Fourier transform (STFT). The Fourier local coefficients are determined for each pixel for four frequency points. The signs of the actual and imaginary components of each coefficient are then calculated using a binary scalar quantity to approximate the phase detail. The resulting 8-bit binary coefficients are then represented as whole by binary coding [46].

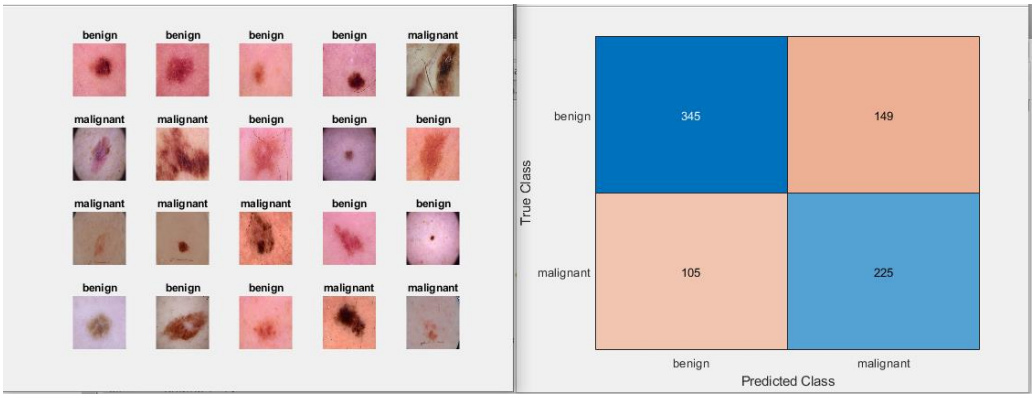

Figure 46: LPQ confusion matrix

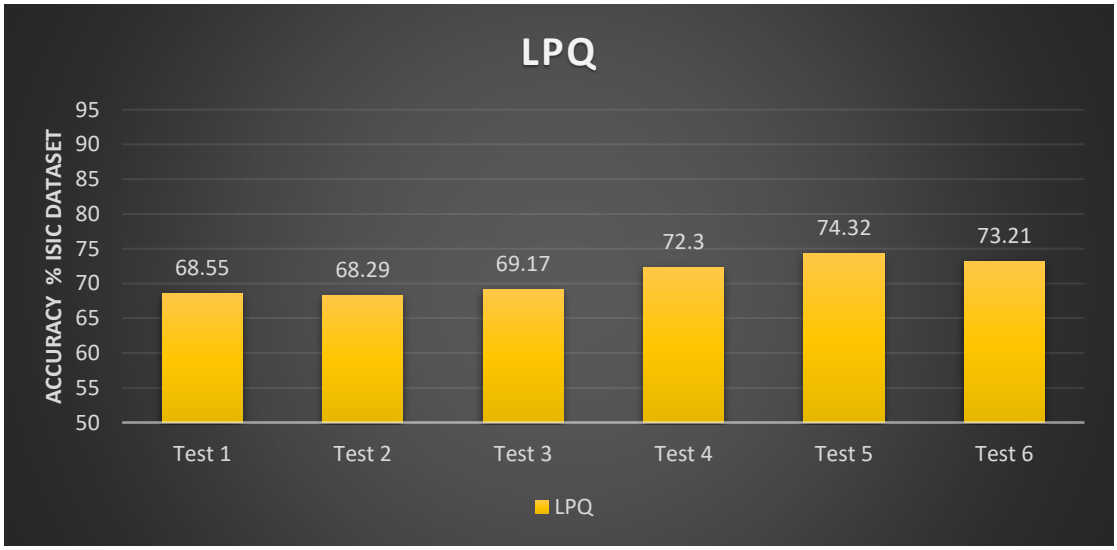

Chart 21: Accuracy of LPQ algorithm using ISIC dataset

CHAPTER SIX

6. Results and Experiments

The effects of all methods and strategies are discussed in this chapter. It is divided into two main sections. In the first section, we experimented with different feature descriptors of skin lesion images, and tested their accuracy on ISIC dataset by using different cell sizes, different Gridhist function values, and changing the size of training data and testing data. We will discuss CNN architectures in the second section.

The major aim of our work is to make awareness about pre-trained neural networks and hand-crafted features extractors for skin cancer classification. These pre-trained neural networks include GoogleNet, densenet201, ResNet-101, ResNet-18, MobileNet-v2 etc.

In order to counteract conventional approaches, some researchers have been able to precisely classify the most important features of the environment, both through entrusting extraction functions and classification tasks to profound learning models. These studies have tried successively, beginning with the exploration of different models, to explore specific learning methods and adaptation, and to improve their approach to deep learning. In this section, we will organize research on various aspects of deep learning in line with their contributions.

In our study, we compared the performance of different CNN models, Resnet, Alexnet, GoogleNet, and ResNet-50, in order to consider CNN suitable for this classification feature. Results demonstrated Resent efficiency and the need for data increase, the completion of all layers, large WSIs instead of small patches, and the combination of various amplification-specific models with Ensemble Learning. The results of our research have examined the performance of another CNN named Inception-v3, which is a GoogleNet version, and our research suggests that this CNN is more successful than previous works.

6.1. Image Level Accuracy

Image level accuracy is standard image accuracy that is taken into account by just giving images to classifier and obtaining the results given as: $ILA = I_{COT} / I_{TOT}$ [36]

Total no of images is illustrated as I_{TOT} and for total no of correctly classified images I_{COT} is mentioned.

Normally, a malignant case is considered positive during cancer diagnosis, while a benign case is viewed as negative. Clinically more significant is a CAD system's response to positive (malignant) events. The first two metrics, ILA and PLA, are based on the typical precision measured for both positive and negative predictions. [60]

The F1score is often referred to as the F1 metric or F score to help highlight the sensitivity of a CAD device to (positive) malignant cases as they demonstrate strong interest in this form of medical diagnosis. The harmonic mean between sensitivity (also known as recall) and precision is the F1-score:

$$F1score = 2 \times Recall \times Precision / Recall + Precision$$

Where:

$$Precision = True\ Positives / True\ Positives + False\ Positives$$

$$Recall = Sensitivity = True\ Positives / True\ Positives + False\ Negatives$$ [1, 60, 61]

Many ISIC CAD systems were developed over the past few years. ISIC data collection for the construction of CAD systems, their strength and vulnerabilities are the key inspiration behind the proposed taxonomy. However, this categorization has allowed us to define and evaluate from both clinical and functional perspectives, as we know, the most suitable reformulation for this issue.

6.2. Achievements and Results using handcrafted features

In our work, we have carried out experiments on multiple classification descriptors such as local binary pattern (LBP) along with its variants, completed local binary pattern, Histogram of Oriented Gradients (HOG) used to extract features from image data, Patterns of Oriented Edge Magnitudes, Binary Statistical Image Features, nearly 27 different hand-crafted feature descriptors for skin lesion detection

In the table below, we used hand-engineered feature extractors and the results are not impressive, but we still learned about the working architectures of different hand-engineered algorithms.

Hand-Crafted Feature Extractors Results											
Classifier Name:	**Test:1**	**T:2**	**T:3**	**T:4**	**T:5**	**T:6**	**T:7**	**T:8**	**T:9**	**T:10**	**Highest Accuracy**
BPPC	63.80	64.21	64.19	x	x	x	x	x	x	x	64.21%
GDP	61.98	58.65	65.02	63.90	x	x	x	x	x	x	65.02%
GDP2	64.81	66.23	70.88	72.90	73.81	74.52	74.42	75.23	75.13	75.13	75.23%
GLTP	66.33	69.06	71.28	73.61	72.50	71.99	x	x	x	x	73.61%
IWBC	68.25	70.37	73.21	72.60	72.80	72.60	x	x	x	x	73.21%
LAP	64.41	66.73	69.36	71.49	71.99	72.90	71.89	x	x	x	72.90%
LBP	64.41	71.79	74.66	73.41	75.27	75.53	77.45	76.34	75.63	75.83	**77.45%**
LDiP	67.85	64.51	68.66	71.39	68.55	x	x	x	x	x	71.39%
LDiPv	72.30	73.00	72.09	70.68	x	x	x	x	x	x	73.00%
LDN	68.25	71.08	73.41	73.41	74.01	75.33	77.25	75.63	76.64	73.81	**77.25%**
LDTP	66.73	68.15	69.06	69.97	68.55	69.87	67.95	x	x	x	69.97%
LFD	63.60	58.54	59.86	62.29	61.38	x	x	x	x	x	63.60%
LGBPHS	64.41	66.63	65.22	x	x	x	x	x	x	x	66.63%
LGDiP	63.60	60.87	x	x	x	x	x	x	x	x	63.60%
LGIP	67.75	70.51	73.71	75.94	75.23	73.91	75.53	73.71	75.94	75.23	75.94%
LGP	61.68	66.16	64.00	68.45	71.59	72.40	69.77	74.52	x	x	74.52%
LGTrP	54.20	54.98	56.55	55.92	57.43	56.72	x	x	x	x	57.43%
LMP	70.68	72.69	74.03	73.51	75.23.	72.80	74.52	73.41	x	x	75.23%
LPQ	68.29	68.29	69.17	72.30	74.32	73.21	x	x	x	x	74.32%
LTeP	70.78	73.90	74.76	74.92	76.14	75.73	75.33	75.43	x	x	76.14%
LTrP	60.57	64.64	59.86	60.97	60.47	57.03	x	x	x	x	64.64%
MBC	68.15	67.22	70.15	70.07	71.99	70.48	x	x	x	x	71.99%
MBP	70.48	72.23	75.36	74.52	74.72	76.04	77.55	76.14	76.24	74.42	**77.55%**
MRELBP	76.85	71.12	72.09	72.60	71.39	72.09	71.99	68.45	70.37	x	76.85%
MTP	70.37	70.56	75.12	75.03	77.55	77.05	76.85	75.23	72.60	72.80	**77.55%**
mWLD	67.64	71.00	73.06	75.94	76.74	75.43	75.33	76.64	74.42	74.92	**76.74%**
PHOG	69.57	71.60	70.02	73.31	71.59	70.88	70.27	69.36	x	x	73.31%

Table 3: Hand Crafted Feature Extractors Result

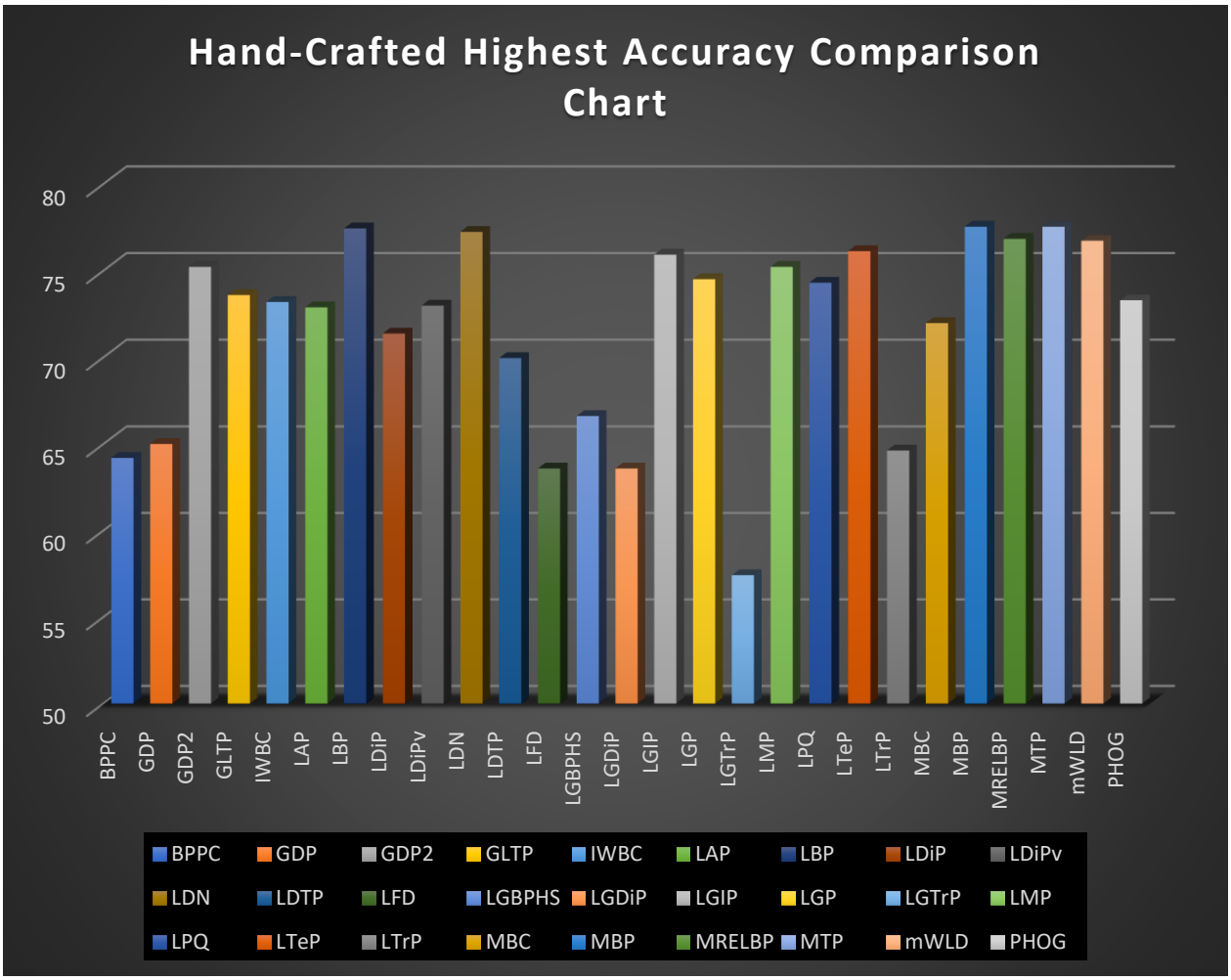

Chart 22: Hand-Engineered Features Accuracy Difference Plot

6.3. Achievements and Results using Deep Learning Models

As we have trained our proposed system on the Multiple Pre-Trained Neural Network Architecture (DCNN) and the findings obtained are also compared with existing techniques. We adopted the same experimental settings suggested by dataset owners and often used by various researchers in their work. We started our project by analyzing the performance of other research works. We measured the performance of the architecture based on the overall accuracy of the predictions.

For the first time, we operated on the InceptionV2, GoogleNet, Xception pre-trained ImageNet for binary scoring. Subsequently, the same fine-tuned ResNet-18 and, respectively, a multi-category rating performance layer were re-tuned. The second phase consisted of two parts, one devoted to benign subclasses and the other to malignant subclasses. Each instance was given its corresponding subclass identification module after defining its main class in the binary classification process.

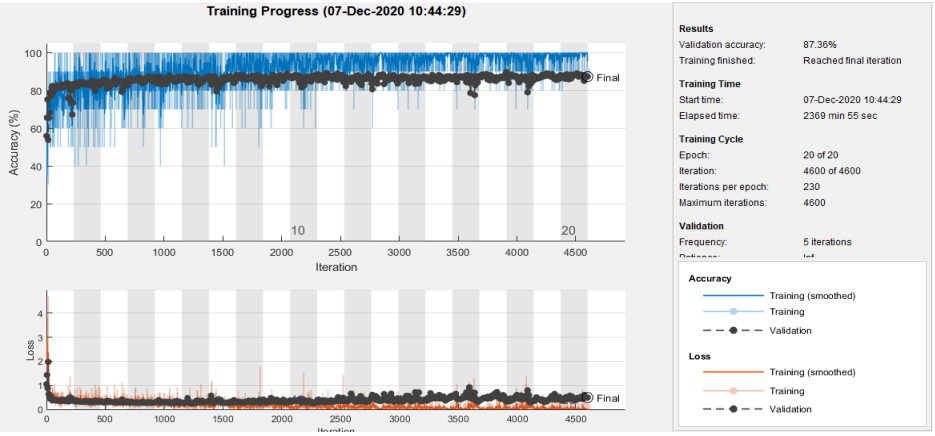

Figure 47: Sample Training Progress of GoogleNet

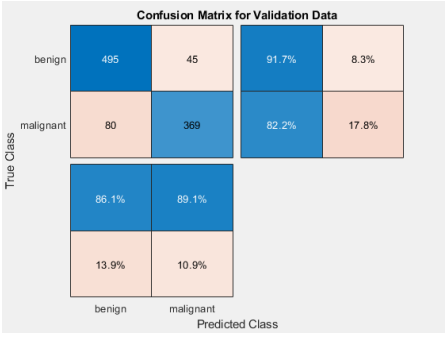

Figure 48: Confusion Matrix of GoogleNet

6.3.1 Inception V3

Google's Inception v3 architecture was re-trained on our dataset by fine-tuning across all layers and replacing top layers with one average pooling, two fully connected and finally the softmax layer allowing to classify 2 diagnostic categories. The size of input images was all resized to (224, 224) to be compatible with this model. Learning rate was set to 0.0001 and epoch were set to 8.

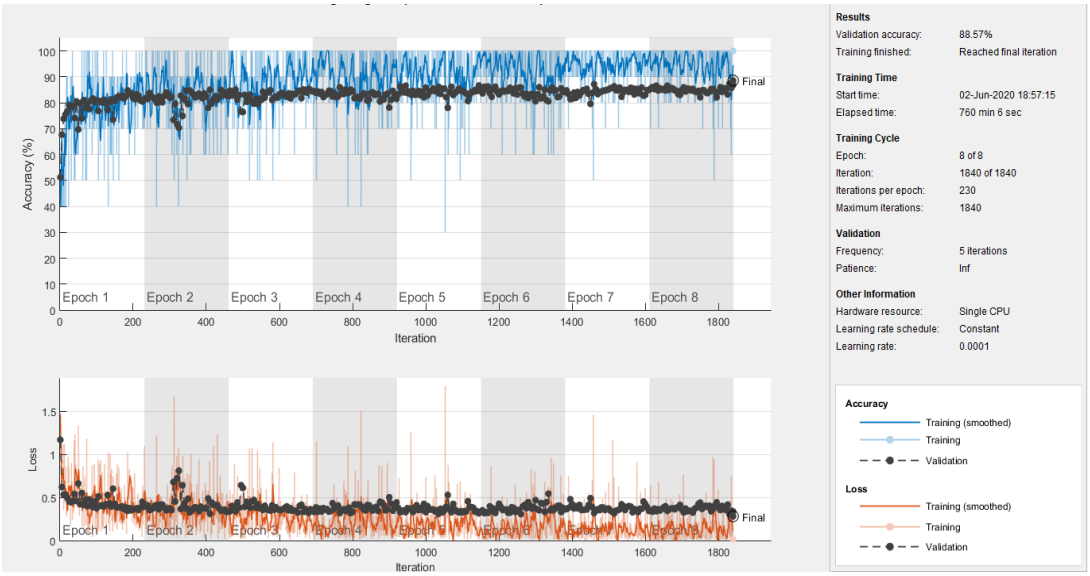

Figure 49: Sample Training Progress of InceptionV3

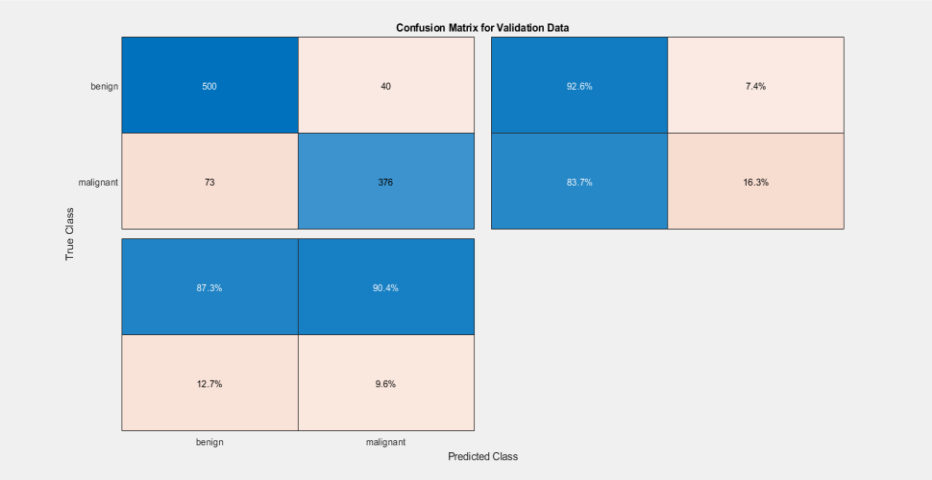

Figure 50: Confusion matrix of InceptionV3

6.3.2 InceptionResNetv2

InceptionResNetv2 architecture was re-trained on our dataset by fine-tuning across all layers and replacing top layers with one global average pooling, one fully connected and finally the SoftMax layer allowing to classify 2 diagnostic categories. The size of input images was all resized to (224, 224) to be compatible with this model. Learning rate was set to 0.0001.

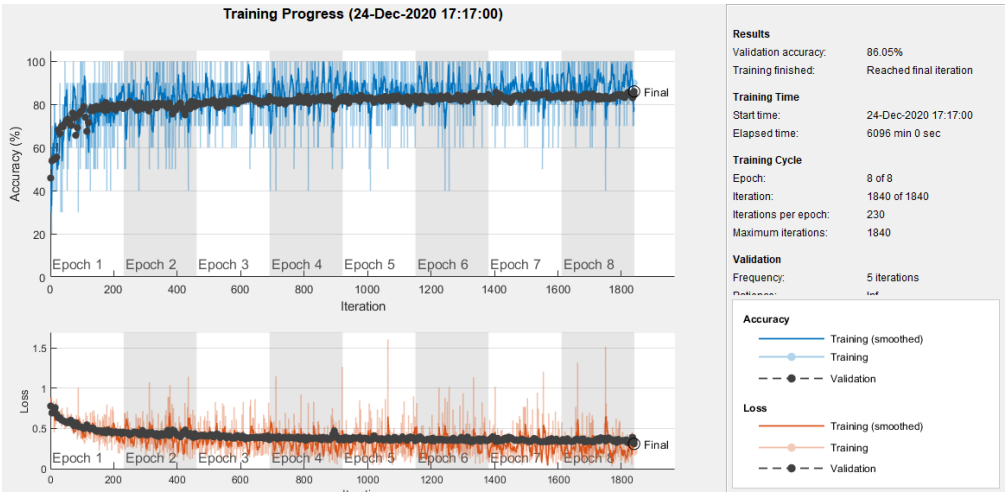

Figure 51: Sample Training Progress of InceptionResnetv2

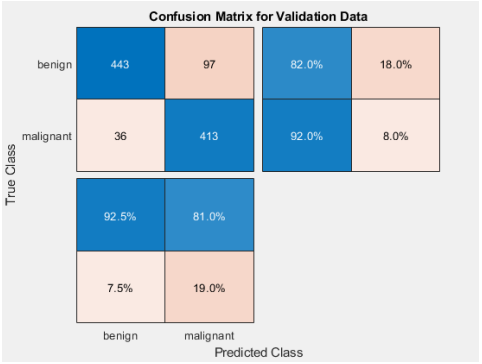

Figure 52: Confusion matrix of InceptionResnetv2

Model	Transfer learning	Train/test	Results
InceptionResnetV2	ImageNet	70%/30%	86.55%
GoogleNet	ImageNet	70%/30%	87.36%
Resnet18	ImageNet	70%/30%	83.61%
InceptionV3	ImageNet	70%/30%	88.57%
Resnet101	ImageNet	70%/30%	**88.98%**
Densnet201	ImageNet	70%/30%	83.62%
InceptionResnetV2	ImageNet	70%/30%	86.55%
mobilenetV2	ImageNet	70%/30%	86.65%
Resnet50	ImageNet	70%/30%	87.56%
	ImageNet	70%/30%	%

Table 4: Comparison of pre-trained weights on results.

In this study we have found that ResNet101 as the backbone gives better results when compared to InceptionV3, InceptionResNetV2, Mobilenet and other pre-trained networks.

	Year	Cancer Types	Accuracy (%)
Pomponiu et al [16]	2016		93.64
Gutman et al [58]			93.1
Haenssle et al [58]	2018		-
Han et al [55]	2018		-
Bi et al [58]	2017		-
Kawahara et al [58]	2018	6	79.5
Lopez et al [58]	2017		81.33
Nasr-Esfahani et al [18]	2016		81
Ramlakhan et al.	2011	2	66.7
Mahbod et al. [58]	2109	2	94.5
Esteva et al. [9]	2017	3	72.1
Ours (Resnet101)		**2**	**88.98**
Ours (InceptionV3)		**2**	**88.57**
Ours (GoogleNet)		**2**	**87.36**

Table 5: Comparison with previous work. [16][58][55][18][9]

The comparison with previous state of the art results is shown in Table 27. Mahbod et al achieves a higher accuracy as they have used the complete dataset. Since our objective

was to show comparable results with limited images, hence our accuracy is good enough considering the number of images on which our model is trained. Bi et al has much better accuracy than ours as they have used an ensemble of CNNs for training, rather than ours which only used a single CNN architecture. Other than that, our work archives state of the art results on skin cancer classification.

CHAPTER SEVEN

7.1. Future Work

In the future, we will focus on enhancing the performance of the CNN model. According to the research carried out by Karpathy A et al, the highest accuracy of CNN can be accomplished by optimizing the data used for model training. An experiment can then be carried out to test the performance of the model by expanding the number of images used for training. We have improved the proposed model by fine-tuning it.

7.2. Conclusion

After study and implementation of deep learning in the field of cancer diseases it is concluded that deep leaning while using Resnet101 provide very competitive results comparing to the other pretrained networks. In our work, we performed ISIC dataset experiments using Deep CNN and also applied a variety of handcrafted features to enhance comprehension of both CNN and handcrafted features. We've seen how we can train and test images of skin cancer on current CNN architectures. In line with the work performed on the same data set, we have obtained enormous and effective results using different methods. Our results differ from those of other research workers, and our results show the performance of CNN and handcrafted features. More analysis can be achieved by integrating CNN with a number of handcrafted feature extractors. Accuracy can be enhanced by eliminating noise and stain from the data set and performing various pre-processing tasks.

References

[1] Vatsala Singh, Ifeoma Nwogu. "Analyzing Skin Lesions in Dermoscopy Images Using Convolutional Neural Networks", 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018.

[2] Cristina Nader Vasconcelos, Bárbara Nader Vasconcelos. "Experiments using deep learning for dermoscopy image analysis", Pattern Recognition Letters, 2020.

[3] J, Deeks JJ, Chuchu N, Ferrante di Ruffano L, Matin RN, Thomson DR, Wong KY, Aldridge RB, Abbott R, Fawzy M, Bayliss SE, Grainge MJ, Takwoingi Y, Davenport C, Godfrey K, Walter FM, Williams HC; Cochrane Skin Cancer Diagnostic Test Accuracy Group. Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults. Cochrane Database Syst Rev. 2018.

[4] Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. Lancet Oncol. 2002.

[5] Moqadam, Sepideh & Grewal, Parvind & Haeri, Zahra & Ingledew, Paris-Ann & Kohli, Kirpal & Golnaraghi, M.. (2018). Cancer detection based on electrical impedance spectroscopy: A clinical study. Journal of Electrical Bioimpedance. 9. 17-23. 10.2478/joeb-2018-0004.

[6] Youssef Filali, Hasnae El Khoukhi, My Abdelouahed Sabri, Ali Yahyaouy, Abdellah Aarab. "Texture Classification of skin lesion using convolutional neural network", 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), 2019.

[7] Md. Fazle Rasul, Md. Fazle Rasul, Nahin Kumar Dey, Nahin Kumar Dey, M. M. A. Hashem, M.M.A. Hashem. "A Comparative Study of Neural Network Architectures for Lesion Segmentation and Melanoma Detection", 2020 IEEE Region 10 Symposium (TENSYMP), 2020.

[8] Chinen, Troy & Ballé, Johannes & Gu, Chunhui & Hwang, Sung & Ioffe, Sergey & Johnston, Nick & Leung, Thomas & Minnen, David & O'Malley, Sean & Rosenberg, Charles & Toderici, George. (2018). Towards A Semantic Perceptual Image Metric. 624-628. 10.1109/ICIP.2018.

[9] Esteva, Andre & Kuprel, Brett & Novoa, Roberto & Ko, Justin & Swetter, Susan & Blau, Helen & Thrun, Sebastian. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature. 542. 10.1038/nature21056

[10] Masood, Ammara & Al-Jumaily, Adel. (2013). Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms. International journal of biomedical imaging. 2013. 323268. 10.1155/2013/323268.

[11] A. Namozov, D. Ergashev and Y. I. Cho, "Adaptive Activation Functions for Skin Lesion Classification Using Deep Neural Networks," 2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS), Toyama, Japan, 2018, pp. 232-235.

[12] V. Singh and I. Nwogu, "Analyzing Skin Lesions in Dermoscopy Images Using Convolutional Neural Networks," 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2018, pp. 4035-4040.

[13] A. Namozov and Y. I. Cho, "Convolutional Neural Network Algorithm with Parameterized Activation Function for Melanoma Classification," 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2018, pp. 417-419.

[14] K. Pai and A. Giridharan, "Convolutional Neural Networks for classifying skin lesions," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 1794-1796.

[15] Kwasigroch, A. Mikołajczyk and M. Grochowski, "Deep neural networks approach to skin lesions classification — A comparative analysis," 2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR), Miedzyzdroje, 2017, pp. 1069-1074.

[16] V. Pomponiu, H. Nejati and N. -. Cheung, "Deepmole: Deep neural networks for skin mole lesion classification," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, 2016, pp. 2623-2627.

[17] H. N. Pham et al., "Lesion Segmentation and Automated Melanoma Detection using Deep Convolutional Neural Networks and XGBoost," 2019 International Conference on System Science and Engineering (ICSSE), Dong Hoi, Vietnam, 2019, pp. 142-147.

[18] E. Nasr-Esfahani et al., "Melanoma detection by analysis of clinical images using convolutional neural network," 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, 2016, pp. 1373-1376.

[19] Enakshi Jana, Ravi Subban, S. Saraswathi. "Research on Skin Cancer Cell Detection Using Image Processing", 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2017.

[20] https://alphabold.com/neural-networks-and-deep-learning-an-overview Sun, 7 June 2020- 16:14:59 UTC

[21] M. AMRANE, S. OUKID, I. GAGAOUA and T. ENSAR, "Skin Cancer Classification Using Machine Learning," in Conference: 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), 2018.

[22] H. Li, S. Zhuang, D.-a. Li, J. Zhao and Y. Ma, "Benign and malignant classification of mammogram images based on deep learning," Biomedical Signal Processing and Control, vol. 51, pp. 347-354, 2019.

[23] Alloghani, Mohamed & Al-Jumeily, Dhiya & Mustafina, Jamila & Hussain, Abir & Aljaaf, Ahmed. (2020). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. 10.1007/978-3-030-22475-2_1.

[24] Bunker, Rory & Thabtah, Fadi. (2017). A Machine Learning Framework for Sport Result Prediction. Applied Computing and Informatics. 15. 10.1016/j.aci.2017.09.005.

[25] F. Ghaznavi, A. Evans and A. M. a. M. Feldman, "Digital Imaging in Pathology: Whole-Slide Imaging and Beyond," Annual Review of Pathology: Mechanisms of Disease, vol. 8, pp. 331-359, 15, 2012.

[26] Mousavi, Seyed. (2018). Recognizing, Distinguishing and Tracking Enemy Army by Missile's RGB-D Sensors, to Decrease Civilian's Casualty, in Missile War.

[27] N. Zeng, Z. Wang, H. Zhang, W. Liu and F. E. Alsaadi, "Deep Belief Networks for Quantitative Analysis of a Gold Immunochromatographic Strip," Cognitive Computation, vol. 8, pp. 684-692, 30 April 2016.

[28] C. Wang, J. Shi, Q. Zhang and S. Ying, "Histopathological image classification with bilinear convolutional neural networks," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, South Korea, 14 September 2017.

[29] K. B. Nahato, K. N. Harichandran and K. Arputharaj, "Knowledge Mining from Clinical Datasets Using Rough Sets and Backpropagation Neural Network," Computational and Mathematical Methods in Medicine, 2015.

[30] https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html (4/08/2020 - 09:55:20 UTC).

[31] Li, Rui & Zhang, Junsong. (2020). Review of computational neuroaesthetics: bridging the gap between neuroaesthetics and computer science. Brain Informatics. 7. 16. 10.1186/s40708-020-00118-w.

[32] J. L. Schonberger, H. Hardmeier, T. Sattler and M. Pollefeys, "Comparative Evaluation of Hand-Crafted and Learned Local Features," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 09 November 2017.

[33] Yang, Meng & Zhang, Lei & Shiu, Simon & Zhang, David. (2012). Monogenic Binary Coding: An Efficient Local Feature Extraction Approach to Face Recognition. Information Forensics and Security, IEEE Transactions on. 7. 1738-1751. 10.1109/TIFS.2012.2217332.

[34] Nguyen, Hieu & Bai, Li & Shen, Linlin. (2009). Local Gabor Binary Pattern Whitened PCA: A Novel Approach for Face Recognition from Single Image Per Person. 269-278. 10.1007/978-3-642-01793-3_28.

[35] M. Yang, L. Zhang, S. C. Shiu and D. Zhang, "Monogenic Binary Coding: An Efficient Local Feature Extraction Approach to Face Recognition," in IEEE Transactions on Information Forensics and Security, vol. 7, no. 6, pp. 1738-1751, Dec. 2012, doi: 10.1109/TIFS.2012.2217332.

[36] Liu L, Lao S, Fieguth PW, Guo Y, Wang X, Pietikäinen M. Median Robust Extended Local Binary Pattern for Texture Classification. IEEE Trans Image Process. 2016 Mar;25(3):1368-81. doi: 10.1109/TIP.2016.2522378. PMID: 26829791.

[37] Z. W. H. Z. W. L. &. F. E. A. Nianyin Zeng, "Deep Belief Networks for Quantitative Analysis of a Gold Immunochromatographic Strip," in Cognitive Computation, 2016.

[38] J. B. Tenenbaum, V. d. Silva and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," Science, vol. 290, no. 5500, pp. 2319-2323, 2000.

[39] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li and A. M.Dobaie, "Facial expression recognition via learning deep sparse autoencoders," Neurocomputing, vol. 273, pp. 643-649, 2018.

[40] Revina, I.Michael & Emmanuel, W.R. Sam. (2018). Face Expression Recognition Using Ldn and Dominant Gradient Local Ternary Pattern Descriptors. Journal of King Saud University - Computer and Information Sciences. 10.1016/j.jksuci.2018.03.015.

[41] Ahmed, Faisal. (2012). Gradient directional pattern: A robust feature descriptor for facial expression recognition. Electronics Letters. 48. 1203-1204. 10.1049/el.2012.1841.

[42] Akhtar, Zahid & Dasgupta, Dipankar. (2019). A Comparative Evaluation of Local Feature Descriptors for DeepFakes Detection. 10.1109/HST47167.2019.9033005.

[43] Jabid, Taskeed & Kabir, Md & Chae, Oksam. (2010). Local Directional Pattern (LDP) A Robust Image Descriptor for Object Recognition. Advanced Video and Signal Based Surveillance, IEEE Conference on. 482-487. 10.1109/AVSS.2010.17.

[44] Ahmed, Faisal & Hossain, Emam. (2013). Automated Facial Expression Recognition Using Gradient-Based Ternary Texture Patterns. Chinese Journal of Engineering. 2013. 1-8. 10.1155/2013/831747.

[45] Mohammad, Tahseen & Ali, Md. Liakot. (2011). Robust facial expression recognition based on Local Monotonic Pattern (LMP). 14th International Conference on Computer and Information Technology, ICCIT 2011. 572-576. 10.1109/ICCITechn.2011.6164854.

[46] Ojansivu, Ville & Heikkilä, Janne. (2008). Blur Insensitive Texture Classification Using Local Phase Quantization. 5099. 236-243. 10.1007/978-3-540-69905-7_27.

[47] M. NEMISSI, H. SALAH and H. SERIDI, "Breast cancer diagnosis using an enhanced Extreme Learning Machine based-Neural Network," in 2018 International Conference on Signal, Image, Vision and their Applications (SIVA), Guelma, Algeria,, 2018.

[48] Yu, Chanki & Yang, Sejung & Kim, Wonoh & Jung, Jinwoong & Chung, Kee-Yang & Lee, Sangwook & Oh, Byungho. (2018). Acral melanoma detection using a convolutional neural network for dermoscopy images. PLOS ONE. 13. e0193321. 10.1371/journal.pone.0193321.

[49] Liang, Qin & Tvete, H & Brinks, H. (2019). Prediction of vessel propulsion power using machine learning on AIS data, ship performance measurements and weather data. Journal of Physics: Conference Series. 1357. 012038. 10.1088/1742-6596/1357/1/012038.

[50] Peng, Jie & Kang, Shuai & Ning, Zhengyuan & Deng, Hangxia & Shen, Jingxian & Xu, Yikai & Zhang, Jing & Zhao, Wei & Li, Xinling & Gong, Wuxing & Huang, Jinhua & Liu, Li. (2019). Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. European Radiology. 30. 1-12. 10.1007/s00330-019-06318-1.

[51] Jeny, Afsana & Junayed, Masum Shah & Atik, Syeda. (2018). PassNet - Country Identification by Classifying Passport Cover Using Deep Convolutional Neural Networks. 1-6. 10.1109/ICCITECHN.2018.8631975.

[52] https://dataconomy.com/2017/04/history-neural-networks/ (22/04/2020 - 19:33:32 UTC).

[53] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision", arXiv: 1512.00567 [cs.CV], 2015

[54] Gao Huang, Zhuang Liu, Laurens van der Maaten and Kilian Q. Weinberger, "Densely Connected Convolutional Networks", arXiv: 1608.06993 [cs.CV], 2016.

[55] Han, Tao & Li, Zhang & Liu, Haixia & Liu, Ping & Ouyang, Quchang & Tang, Yuling & Hu, Zheyu & Li, Qiang. (2018). Optimize Transfer Learning for Lung Diseases in Bronchoscopy Using a New Concept: Sequential Fine-Tuning. IEEE Journal of Translational Engineering in Health and Medicine. 10.1109/JTEHM.2018.2865787.

[56] Noh, Kyoung & Choi, Jiho & Hong, Jin & Park, Kang. (2020). Finger-Vein Recognition Based on Densely Connected Convolutional Network Using Score-Level Fusion With Shape and Texture Images. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2996646.

[57] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks", arXiv: 1801.04381 [cs.CV], 2018.

[58] Sagar, Abhinav & Jacob, Dheeba. (2020). Convolutional Neural Networks for Classifying Melanoma Images. 10.13140/RG.2.2.34034.53446/1.

[59] Ponzio, Francesco & Macii, Enrico & Ficarra, Elisa & Di Cataldo, Santa. (2018). Colorectal Cancer Classification using Deep Convolutional Networks - An Experimental Study. 58-66. 10.5220/0006643100580066.

[60] https://www.isic-archive.com/#!/topWithHeader/tightContentTop/about/isicArchive

[61] Pollastri, Federico & Bolelli, Federico & Paredes, Roberto & Grana, Costantino. (2020). Augmenting data with GANs to segment melanoma skin lesions. Multimedia Tools and Applications. 79. 1-18. 10.1007/s11042-019-7717-y.