CHAPTER 1 INTRODUCTION

In today's real time scenario of daily life, computer vision methodology has attracted researchers due to its nature of providing efficient information for better visual and experimental analysis. In the computer vision approach, image classification is also a promising technique which is used for various applications such as pattern recognition, remote sensing application, medical image processing etc. It is a process of pixel sorting from image and accumulating into individual classes. For classification, various methods have been developed to classify and recognize the image class efficiently.

1.1 IDENTIFYING THE DISEASE:

Dermatology remains the most uncertain and complicated branch of science because of its complication in the procedures involved in diagnosis of diseases related to hair, skin, nails. The variation in these diseases can be seen because of many environmental, geographical factor variations.Human skin is considered the most uncertain and troublesome terrains due to the existence of hair, its deviations in tone and other mitigating factors. The skin disease diagnosis includes series of pathological laboratory tests for the identification of the correct disease.

1.2 HISTORY OF THE DISEASE:

For the past ten years these diseases have been the matter of concern as their sudden arrival and their complexities have increased the life risks.1 These Skin abnormalities are very infectious and need to be treated at earlier stages to avoid it from spreading. Total wellbeing including physical and mental health is also adversely affected. Many of these skin abnormalities are very fatal particularly if not treated at an initial stage.

1.3 SCOPE

The project will be able to predict the type of skin disease by giving or feeding data to the model to which it will use CNN algorithm and predict using the 4 classes and give a label to the disease.Our project will be useful for people who are suffering from skin disease and who do not have convenience to visit a doctor.Just with internet connection a person can check his/her disease by clicking a photo and uploading so the model will predict and label the disease for them.

CHAPTER 2 PROBLEM DEFINITION

Currently skin disease identification is performed by medical professionals; humans must often search through many skin disease images before finding the desired type of disease. This process of manual recognition is slow and possesses a degree of subjectivity which is hard to be quantified.Skin disease diagnosis includes a series of pathological laboratory tests for the identification of the correct disease,In rural areas its difficult for people to have access to dermatology clinics. In some situations if some people have some skin condition and they want immediate assistance they can't go to the doctor instantly. Skin diseases have a serious impact on people's life and health. Current research proposes an efficient approach to identify singular types of skin diseases. It is necessary to develop automatic methods in order to increase the accuracy of diagnosis for multitype skin diseases.

CHAPTER 3 LITERATURE SURVEY

Comparison of skin disease prediction by feature selection using ensemble data mining techniques.

Background: Skin disease is a major problem among people worldwide. Different machine Learning techniques can be applied to identify classes of skin disease. Herein, we have applied machine learning algorithms to categorize classes of skin disease using ensemble techniques,and then a feature selection method is utilized to compare the results obtained.

Method: In the proposed study, we present a new method which applies six different data mining classification techniques, and then develop an ensemble approach using Bagging, AdaBoost and Gradient Boosting classifier techniques to predict classes of skin disease. Furthermore, a feature importance method is utilized to select the most salient 15 features which will play a major role in prediction. A subset of the original dataset is obtained after selecting the 15 features, to compare the results of six machine learning techniques, and an ensemble approach is applied to the entire dataset.

Results: The ensemble method is compared with the subset obtained from the feature selection method. The outcome shows that the dermatological prediction accuracy of the test dataset is increased as compared to the use of an individual classifier, and improved accuracy is obtained as compared with the feature selection subset method.

Conclusion: The ensemble method and feature selection applied to dermatology datasets yields a better performance as compared to individual classifier algorithms. The ensemble method provides a more accurate and effective skin disease prediction.

Skin Lesion Analysis Toward Melanoma Detection: A Challenge At The 2017 International Symposium On Biomedical Imaging (Isbi), Hosted By The International Skin Imaging Collaboration (Isic)

This article describes the design, implementation, and results of the latest installment of the dermoscopic image analysis benchmark challenge. The goal is to support research and development of algorithms for automated diagnosis of melanoma, the most lethal skin cancer. The challenge was divided into 3 tasks: lesion segmentation, feature detection, and disease classification. Participation involved 593 registrations, 81 pre-submissions, 46 finalized submissions (including a 4-page manuscript), and approximately 50 attendees, making this the largest standardized and comparative study in this field to date. While the official challenge duration and ranking of participants has concluded, the dataset snapshots remain available for further research and development.

CHAPTER 4 PROJECT DESCRIPTION

PROPOSED DESIGN

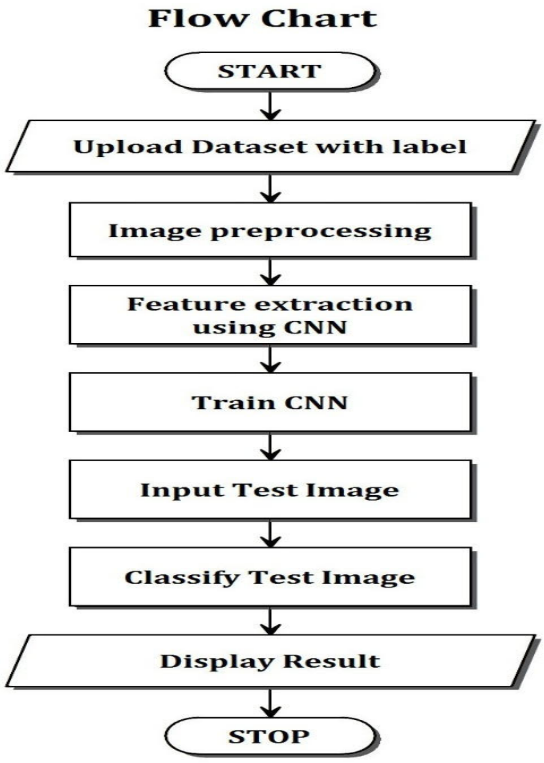


Fig:4.1

In this flowchart first we upload datasets with label to the model,the image preprocessing happens for the images uploaded,using cnn we extract some features for the image process,and then we train the cnn model .we give test image as input to check for the image classification and then the model classifies the image and displays the result .

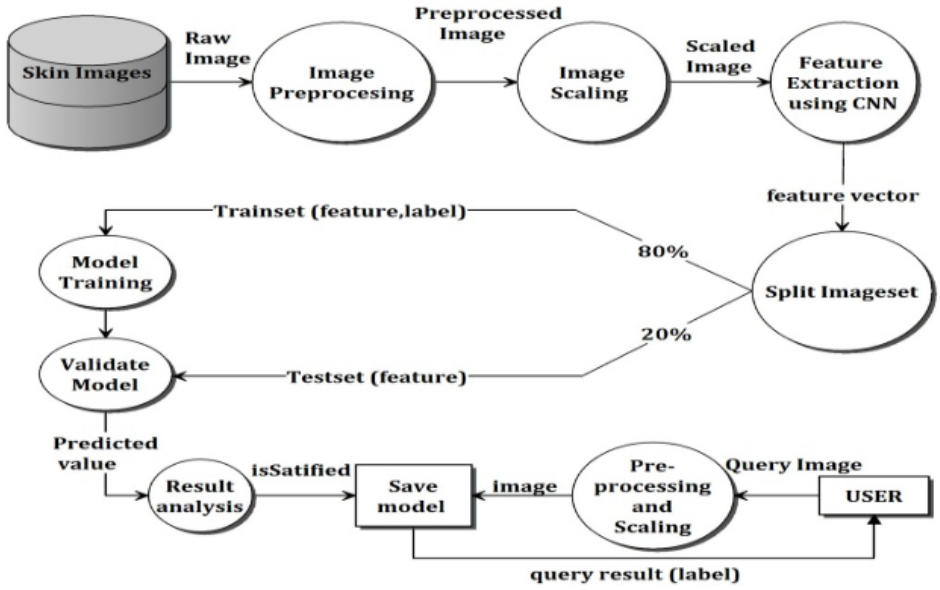


Fig:4.2 Data flow diagram

1. A data flow diagram (DFD) is a graphic representation of the "flow" of data through an information system. A data flow diagram can also be used for the visualization of data processing (structured design).

2. Squares representing external entities, which are sources and destinations of information entering and leaving the system.

3. Rounded rectangles representing processes, in other methodologies, may be called Activities,Actions,Procedures,Subsystems etc. which take data as input, do processing to it, and output it.

4. Arrows representing the data flows, which can either be electronic data or physical items. It is impossible for data to flow from data store to data store except via a process, and external entities are not allowed to access data stores directly.

5. The flat three-sided rectangle representing data stores should both receive information for storing and provide it for further processing.

CHAPTER 5 REQUIREMENTS

5.1 Functional requirements

A function of a software system is defined in functional requirements and the behavior of the system is evaluated when presented with specific inputs or conditions which may include calculations, data manipulation and processing and other specific functionality. The functional requirements of the project are one of the most important aspects in terms of the entire mechanism of modules.Once our model is built then it should be able to classify the skin disease image.

5.2 Nonfunctional requirements

Nonfunctional requirements describe how a system must behave and establish constraints of its functionality. This type of requirements is also known as the system's quality attributes.Attributes such as performance, security, usability, compatibility are not the feature of the system, they are a required characteristic. They are properties that emerge from the whole arrangement and hence we can't compose a particular line of code to execute them.Any attributes required by the customer are described by the specification. We must include only those requirements that are appropriate for our project.

5.3 Software requirements

The following hardware and software are required for the development and deployment of the system.

- **Operating system** : Windows 10
- **Coding Language** : Python
- **Software** : Anaconda
- **IDE** : Jupyter Notebook

5.4 Hardware Requirements:

- System : Intel I3 & above 2.4 GHz.
- Hard Disk : 1 TB.
- Ram : 8 GB

CHAPTER 6 METHODOLOGY

It will cover the detailed explanation of methodology that is being used to make this project complete and working well. Many methodology or findings from this field mainly generated into journals for others to take advantage of and improve as upcoming studies. The Method is used to achieve the objective of the project that will accomplish a perfect result. Inorder to evaluate this project, the methodology based on System Development Life Cycle(SDLC), generally three major step, which is planning, implementing and analysis.

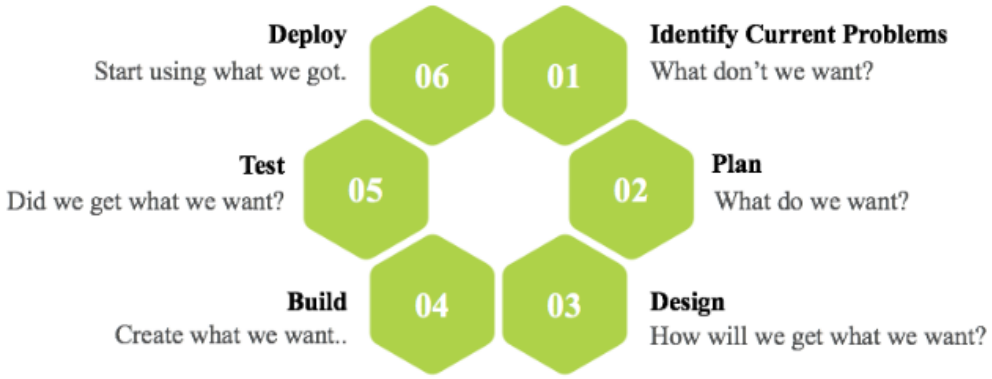


Fig:6.1 Software Development life cycle

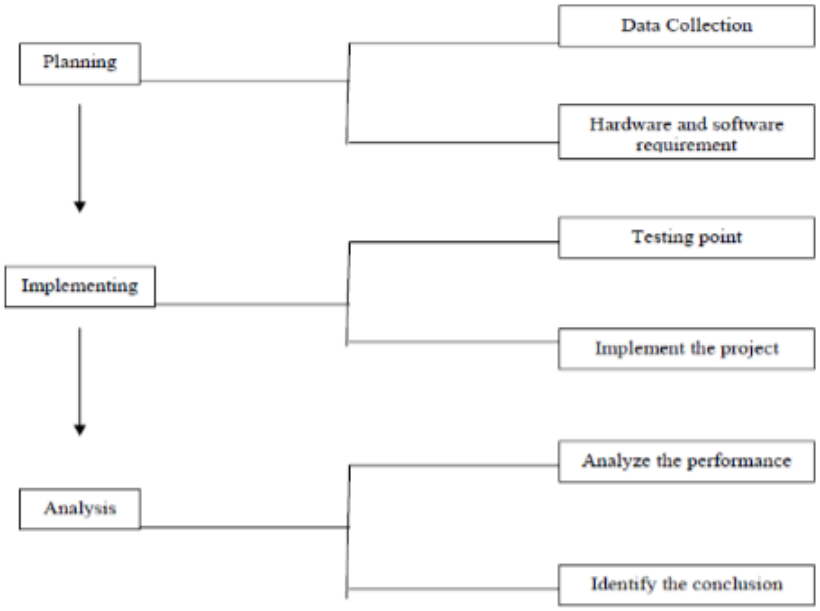


Fig:6.2 Steps of Methodology

Planning:

To identify all the information and requirements such as hardware and software,planning must be done in the proper manner. The planning phase has two main elements namely data collection and the requirements of hardware and software.

Data collection:

Machine learning needs two things to work, data (lots of it) and models. When Acquiring the data, be sure to have enough features (aspect of data that can help for prediction, like the surface of the house to predict its price) populated to train correctly your learning model. In general, the more data you have the better so make sure to come with enough rows.The primary data collected from the online sources remains in the raw form of statements, digits and qualitative terms. The raw data contains error, omissions and inconsistencies.It requires corrections after careful scrutinizing the completed questionnaires.

The following steps are involved in the processing of primary data. A huge volume of raw data collected through field survey needs to be grouped for similar details of individualresponses.Data Preprocessing is a technique that is used to convert the raw data into a clean data set.In other words, whenever the data is gathered from different sources it is collected in rawformat which is not feasible for the analysis.Therefore, certain steps are executed to convert the data into a small clean data set. This Technique is performed before the execution of Iterative Analysis. The set of steps is known as Data Preprocessing. It includes -

- Data Cleaning
- Data Integration
- Data Transformation
- Data Reduction

Data Preprocessing is necessary because of the presence of unformatted real-world data.

Mostly real-world data is composed of -

Inaccurate data (missing data) - There are many reasons for missing data such as

data is not continuously collected, a mistake in data entry, technical problems with biometrics and much more.

The presence of noisy data (erroneous data and outliers) - The reasons for the existence of noisy data could be a technological problem of a gadget that gathers data, human mistakes during data entry and much more.

Inconsistent data - The presence of inconsistencies are due to the reasons such that existence of duplication within data, human data entry, containing mistakes in codes or names, i.e., violation of data constraints and much more.

Implementing:

In this work, a business intelligent model has been developed, to classify different animals,based on a specific business structure dealing with Animal classification using a suitable machine learning technique. The model was evaluated by a scientific approach to measure accuracy. We are using Convolutional Neural Network (CNN) to build our model.

Convolutional Neural Network

A convolutional neural network (CNN) is a special architecture of artificial neural networks,proposed by Yann LeCun in 1988. CNN uses some features of the visual cortex. One of the most popular uses of this architecture is image classification. For example Facebook uses CNN for automatic tagging algorithms, Amazon—for generating product recommendations and Google for search through among users' photos. Let us consider the use of CNN for image classification in more detail. The main task of image classification is acceptance of the input image and the following definition of its class. This is a skill that people learn from their birth and are able to easily determine that the image in the picture is an elephant. But the computer sees the pictures quite differently:

Instead of the image, the computer sees an array of pixels. For example, if image size is 300 x300. In this case, the size of the array will be 300x300x3. Where 300 is width, next 300 is height and 3 is RGB channel values. The computer is assigned a value from 0 to 255 to each of these numbers. This value describes the intensity of the pixel at each point.To solve this problem the computer looks for the characteristics of the base level. In human understanding such characteristics are for example the trunk or large ears. For the computer, these characteristics are boundaries or curvatures. And then through the groups of convolutional layers the computer constructs more abstract concepts.

In more detail: the image is passed through a series of convolutional, nonlinear, pooling layers and fully connected layers, and then generates the output.

The Convolution layer is always the first. The image (matrix with pixel values) is entered into it. Imagine that the reading of the input matrix begins at the top left of the image. Next the software selects a smaller matrix there, which is called a filter (or neuron, or core). Then the filter produces convolution, i.e. moves along the input image. The filter's task is to multiply its values by the original pixel values. All these multiplications are summed up. One number is obtained in the end. Since the filter has read the image only in the upper left corner, it moves further and further right by 1 unit performing a similar operation. After passing the filter across all positions, a matrix is obtained, but smaller than an input matrix.

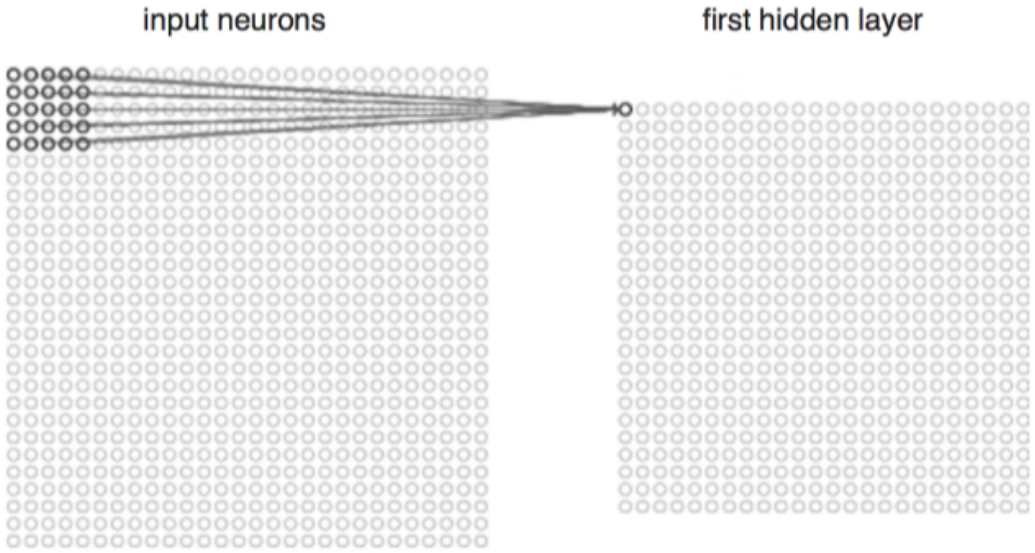


Fig:6.3 The Convolution layer

This operation, from a human perspective, is analogous to identifying boundaries and simple colors on the image. But in order to recognize the properties of a higher level such as the trunk or large ears the whole network is needed.

The network will consist of several convolutional networks mixed with nonlinear and pooling layers. When the image passes through one convolution layer, the output of the first layer becomes the input for the second layer. And this happens with every further convolutional layer.

The nonlinear layer is added after each convolution operation. It has an activation function, which brings nonlinear properties. Without this property a network would not be sufficiently intense and will not be able to model the response variable (as a class label).

The pooling layer follows the nonlinear layer. It works with width and height of the image and performs a down sampling operation on them. As a result the image volume is reduced. This means that if some features (as for example boundaries) have already been identified in the previous convolution operation, than a detailed image is no longer needed for further processing, and it is compressed to less detailed pictures.

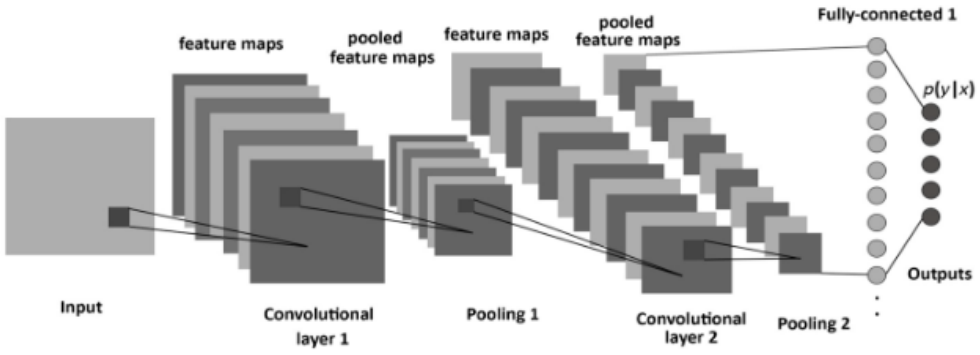


Fig:6.4 The Convolution layer

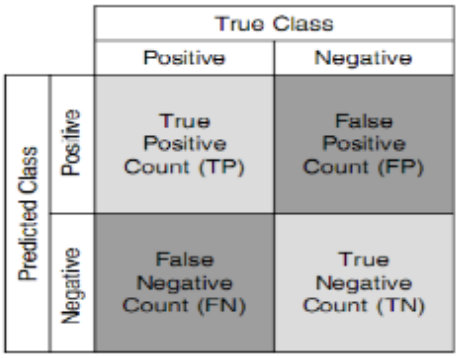
After completion of a series of convolutional, nonlinear and pooling layers, it is necessary to attach a fully connected layer. This layer takes the output information from convolutional networks. Attaching a fully connected layer to the end of the network results in a dimensional vector, where N is the amount of classes from which the model selects the desired class.

Analysis:

In this final phase, we will test our classification model on our prepared image dataset and also measure the performance on our dataset. To evaluate the performance of our created classification and make it comparable to current approaches, we use accuracy to measure the effectiveness of classifiers.

After model building, knowing the power of model prediction on a new instance, is very important issue. Once a predictive model is developed using the historical data, one would be curious as to how the model will perform on the data that it has not seen during the model building process. One might even try multiple model types for the same prediction problem, and then, would like to know which model is the one to use for the real-world decision making situation, simply by comparing them on their prediction performance (e.g., accuracy).

To measure the performance of a predictor, there are commonly used performance metrics, such as accuracy, recall etc. First, the most commonly used performance metrics will be described, and then some famous estimation methodologies are explained and compared to each other. "Performance Metrics for Predictive Modeling In classification problems, the primary source of performance measurements is a coincidence matrix (classification matrix or a contingency table)". Above figure shows a coincidence matrix for a two-class classification problem. The equations of the most commonly used metrics that can be calculated from the coincidence matrix.



$$\text{True Positive Rate} = \frac{TP}{TP + FN}$$

$$\text{True Negative Rate} = \frac{TN}{TN + FP}$$

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

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Fig:6.5 Confusion matrix and formulae

As seen in the above figure, the numbers along the diagonal from upper-left to lower-right represent the correct decisions made, and the numbers outside this diagonal represent the errors. "The true positive rate (also called hit rate or recall) of a classifier is estimated by dividing the correctly classified positives (the true positive count) by the total positive count. The false positive rate (also called a false alarm rate) of the classifier is estimated by dividing the incorrectly classified negatives (the false negative count) by the total negatives. The overall accuracy of a classifier is estimated by dividing the total correctly classified positives and negatives by the total number of samples.

CHAPTER 7 EXPERIMENTATION

- we will test our classification model on our prepared image dataset and also measure the performance on our dataset.
- To evaluate the performance of our created classification and make it comparable to current approaches, we use accuracy to measure the effectiveness of classifiers.
- Once a predictive model is developed using the historical data, one would be curious as to how the model will perform on the data that it has not seen during the model building process.
- To measure the performance of a predictor, there are commonly used performance metrics,such as accuracy, recall etc.
- Preparing test cases, test data, Preparing test procedure, Preparing test scenario, Writing test script.
- In this phase we execute the documents that are prepared in the test development phase.
- Once executed, documents will get results either pass or fail. We need to analyze the results during this phase.
- A result is the final consequence of actions or events expressed qualitatively or quantitatively. Performance analysis is an operational analysis, is a set of basic quantitative relationships between the performance quantities.

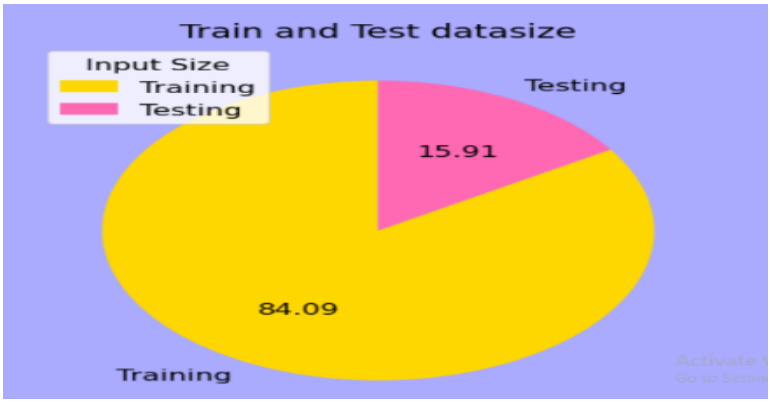


Fig:7 pie chart

CHAPTER 8 TESTING AND RESULTS

Test ID	Test Input	Excepted Result	Actual Result	Remarks
T_01	upload dataset	Uploaded data has to be stored in the work environment	Uploaded data stored in work environment.	Pass
T_02	Feature Extraction	During this process from the images features has to extracted and stored in data frames.	Features are extracted and stored in data frames	Pass
T_03	Labeling process	Based on the image category extracted features has to be labeled.	Extracted feature records are labeled propertly.	Pass
T_04	Data splitting	During this process the data set has to split into training and test data set	The data set spited into training and test data set	Pass
T_05	Training process	This process has to read all the training dataset and create valid data model	This process system read all the training dataset and created valid data model	Pass
T_06	Testing process	This process has to read test data and pass it to validation model and display its classification	This process read test data feature and pass it to validation model and display its classification	Pass

Table no:(8)

Fig:8.1 Login page

Fig:8.2 Home page

Fig : 8.3 Result

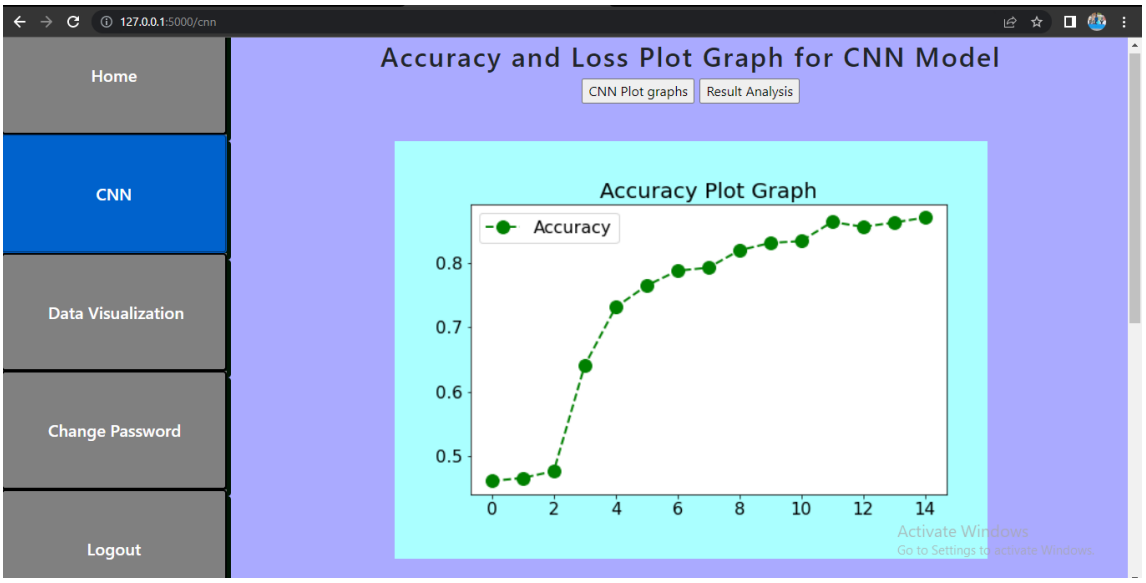


Fig :8.4 Accuracy plot graph

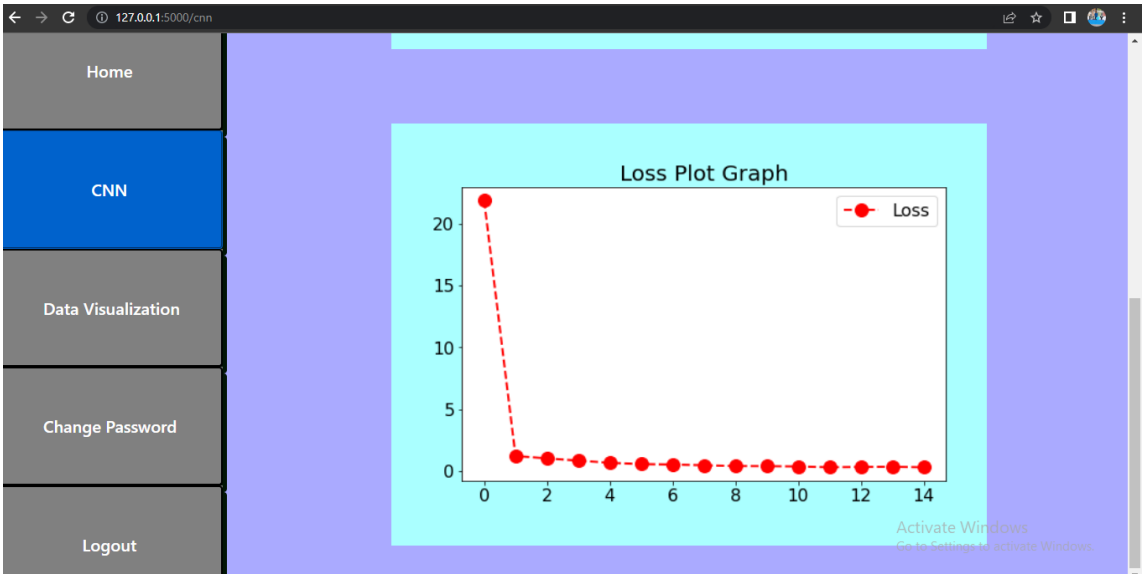


Fig : 8.5 Loss plot graph

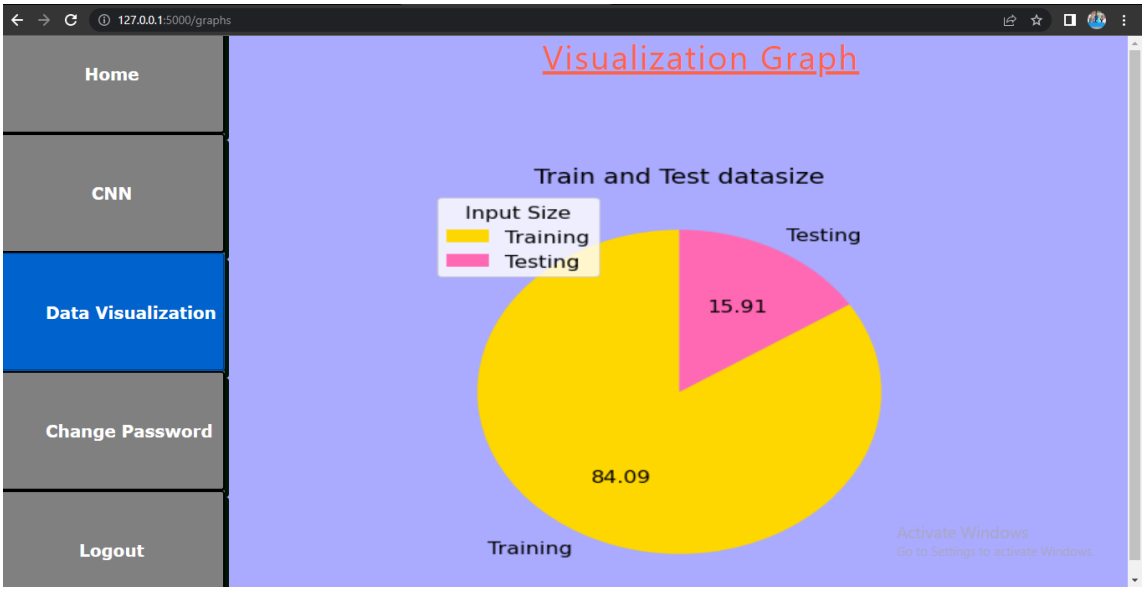


Fig :8.6 Visualization graph

CHAPTER 9:Conclusion And Future Work

The proposed system was implemented in anaconda jupyter notebook on a Windows 10 64-bit operating system with Intel(R) Core(TM) i5-6100U CPU @ 2.30GHz processor. SkinDiseases being extremely common must be detected at the earliest stage. The proposed system in this paper provides a feasible solution for skin disease detection providing up to more than 80% efficiency in CNN. This paper gives the description, result analysis and comparison of the efficiency between the three transforms CNN.Bio-impedance measurement method is for analysis of skin diseases is used in diagnosis of early stage skin diseases like Acne, benign, Dermatitis, Eczema,From above results we conclude that normal skin magnitude has more value than disease skin and we find affected skin. By using this measurement we easily diagnose and compare affected skin with normal skin of any disease .From implementation of the proposed system, we can conclude that the parallel combination of the three transforms provides the maximum and efficient detection. It can be used as a basic prototype to pave the way for faster diagnosis of skin diseases.To achieve higher accuracy, all the three transforms were combined and then disease detection was performed.

REFERENCES

[1] M. DePietro and V. Hiugeria,"Skin infection: types, causes, and treatment",Healthline,2017.[Online].Available:http://www.healthline.com/health/skin-infection. [Accessed: 17- Apr- 2017].

[2] D. Dimri, V. Reddy B and A. Kumar Singh, "Profile of skin disorders inunreachedhillyareasofnorthIndia&,Hindawi,2017.Available:https://www.hindawi.com/journals/drp/2016/8608534/. [Accessed: 15- Apr- 2017].

[3]BioSpectrumBureau,"SkindiseasestogrowinIndiaby2015:Report", Biospectrum,2014.[Online].Available:http://www.biospectrumindia.com/news/73/8437/skin-diseases-to-grow-in-india-by-2015-report.html. [Accessed: 07- Oct- 2016].

[4] E H. Page, MD, Assistant Clinical Professor of Dermatology; Physician, Harvard MedicalSchool; Lahey Hospital and Medical Center "Diagnosis of skindisorders&,MSDManualConsumerVersion,2017.[Online].Available:http://www.merckmanuals.com/home/skindisorders/biology-of-the-skin/diagnosis-of-skin-disorders. [Accessed: 07- Oct- 2016].

[5]M.Rodrishy;guez,Themedianfilterproblems,Tracer.lcc.uma.es.[Online].Available:http://tracer.lcc.uma.es/problem/mfp/mfp.html. [Accessed: 07- Oct- 2016].

[6]P.Miami,"10mostcommonskindiseases&,PositiveMed,2014.[Online].Available:http://positivemed.com/2014/04/22/10-commonskin-diseases/. [Accessed: 04- Aug- 2016].

[7]Discretecosinetransform-MATLAB&Simulink,In.mathworks.com,2016.[Online].Available:https://in.mathworks.com/help/images/discrete-cosine-transform.html [Accessed: 21- Oct- 2016].

[8] "The discrete cosine transform (DCT)", Users.cs.cf.ac.uk, 2017.[Online].Available:https://users.cs.cf.ac.uk/Dave.Marshall/Multimedia/node231.html.[Accessed: 18- Apr- 2017].

[9] M. Sifuzzaman, M.R. Islam and M.Z. Ali, "Application of wavelet transform anditsadvantagescomparedtofouriertransform",JournalofPhysical Sciences, Vol. 13, 2009.

[10]W.K.Saa'd,"Methodfordetectionanddiagnosisoftheareaofskindisease based on colour by wavelet transform and artificial neural network", Al-Qadisiya Journal for Engineering Sciences, Vol. 2,No.4, 2009

[11]Indira,J.SupriyaP,Detectionanalysisofskincancerusingwaveletechniques&,InternationalJournalofComputerScienceandInformation Technologies, Vol. 2 (5), 2011.