Skin Disease Detection Using ML Techniques

Jammala Prem Ajay

Dept of Computer Science, SV University, Tirupati

Abstract— *Skin is the human body's exterior integument. Human skin pigmentation varies from person to person, and skin types include dry, oily, and mixed. The human skin's diversity offers bacteria and other microbes with a diverse home. Melanocytes in the human skin create melanin, which can absorb harmful UV radiation from the sun, causing skin damage and cancer. In most third-world societies, the requisite technologies for early identification of many diseases are still unavailable. If acne, dermatomyositis, candidiasis, cellulitis, Scleroderma, chicken pox, ringworm, eczema, psoriasis, and other skin illnesses are left untreated in their early stages, they can lead to a variety of health issues, including death. The technique of image segmentation aids in the diagnosis of various skin disorders. The goal of this research is to use image processing techniques to diagnose the skin illness from a given image set. Deblurring and noise reduction were performed on the captured image set before it was processed. The disease is detected at the output for a corresponding input image based on the definite pattern (pertaining to a distinct disease) present in the processed image. If acne, dermatomyositis, candidiasis, cellulitis, Scleroderma, chicken pox, ringworm, eczema, psoriasis, Melonama, and other skin illnesses are left untreated in their early stages, they can lead to a variety of health consequences and even death. The technique of image segmentation aids in the diagnosis of various skin disorders. They have taken two classes here. To classify the condition, there are two types of skin: normal and abnormal. Melonama and Acne photos will be processed in the system in the abnormal section*.

I. INTRODUCTION

In India, skin problems affect roughly 10% to 12% of the population. The skin protects the body while also receiving sensory input from the surrounding environment. It is the largest organ in the human body, with seven layers of ectodermic tissue protecting bones, muscles, and internal organs. Poor hygiene, higher levels of pollution, global warming, and dangerous UV radiation are all factors that contribute to skin problems. A one percent ozone depletion can result in a two to three percent increase in cancers. In India, photosensitive and infectious skin disorders are rather widespread. It is critical to treat them as soon as possible, or else complications may arise that impact not only the skin, but also the mental health and life of the individual in the system. For the following reasons, people require effective solutions: Because India's population is rapidly growing, it is critical to provide effective care to each individual at a faster rate. Skin disorders are analogous to diseases such as tuberculosis, AIDS, and others that are fatal if not treated promptly. The cost of treating even a minor skin illness limits an individual's treatment options. As a result, cost-effective yet efficient methods of skin disease identification are critical. Dermatology is the branch of medicine that deals with the diagnosis and treatment of skin, hair, and nail illnesses, and a dermatologist is a qualified medical expert in this subject. In today's world, computerised systems are used in practically all sectors, including medical and other fields, to replace manual traditional equipment with automated technology. With so many different forms of skin disorders, all modern researchers, particularly in medical science, are looking for a solution that may successfully assist doctors in recognising a disease at its early stage of diagnosis without wasting a lot of time utilising modern technology. Digital image processing comes in handy in this situation. Digital image processing is the process of enhancing a digital image by extracting relevant information with the use of computer algorithms. The three main phases of image processing are: importing the image using image acquisition tools; analysing and manipulating the image; and producing a report or output based on the image analysis. The splitting of an image into discrete parts containing pixels with similar properties is known as image segmentation. It facilitates the examination and interpretation of images in a more straightforward manner. Using several segmentation techniques, the researchers hope to successfully detect skin illnesses such as ringworm, eczema, chicken pox, and psoriasis. The accuracy of feature measurement aids in determining the picture segmentation outcome. This paper will assist us in determining the most appropriate image processing strategy for detecting the aforementioned skin diseases: (1) Data collection (2) Pre-processing (3) Segmentation (4) Feature extraction and classification The methods described above are applied to photographs of each of the skin disorders using Matlab Software and Toolbox, and the differences in the outcomes of each dataset are observed.

1.1 Image Processing

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too.

Image processing basically includes the following three steps:

- Importing the image via image acquisition tools;
- Analysing and manipulating the image;
- Output in which result can be altered image or report that is based on image analysis.

There are two types of methods used for image processing namely, analogue and digital image processing. Analogue image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in manipulation of the digital images by using computers. The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction.

Digital image processing is the use of a digital computer to process digital images through an algorithm. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modeled in the form of multidimensional systems. The generation and development of digital image processing are mainly affected by three factors: first, the development of computers; second, the development of mathematics (especially the creation and improvement of discrete mathematics theory); third, the demand for a wide range of applications in environment, agriculture, military, industry and medical science has increased. Digital Image Processing means processing digital image by means of a digital computer. We can also say that it is a use of computer algorithms, in order to get enhanced image either to extract some useful information. A image is defined as a two-dimensional function, **F(x,y),** where x and y are spatial coordinates, and the amplitude of **F** at any pair of coordinates (x,y) is called the **intensity** of that image at that point. When x,y, and amplitude values of **F** are finite, we call it a **digital image**. In other words, an image can be defined by a two-dimensional array specifically arranged in rows and columns. Digital Image is composed of a finite number of elements, each of which elements have a particular value at a particular location. These elements are referred to as picture elements, image elements, and pixels. A Pixel is most widely used to denote the elements of a Digital Image.

Types of an image

- **BINARY IMAGE–** The binary image as its name suggests, contain only two pixel elements i.e 0 & 1,where 0 refers to black and 1 refers to white. This image is also known as Monochrome.
- **BLACK AND WHITE IMAGE–** The image which consist of only black and white color is called BLACK AND WHITE IMAGE.
- **8 bit COLOR FORMAT–** It is the most famous image format. It has 256 different shades of colors in it and commonly known as Grayscale Image. In this format, 0 stands for Black, and 255 stands for white, and 127 stands for gray.
- **16 bit COLOR FORMAT–** It is a color image format. It has 65,536 different colors in it. It is also known as High Color Format. In this format the distribution of color is not as same as Gray scale image.

1.2 Machine Learning

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values. Recommendation engines are a common use case for machine learning. Other popular uses include fraud detection, spam filtering, malware threat detection, business process automation (BPA) and predictive maintenance. Machine learning is important because it gives enterprises a view of trends in customer behavior and business operational

patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks. A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning. Some implementations of machine learning use data and neural networks in a way that mimics the working of a biological brain. In its application across business problems, machine learning is also referred to as predictive analytics.

1.3 Types of Machine Learning

- **Supervised learning:** In this type of machine learning, data scientists supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is specified.
- **Unsupervised learning:** This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through data sets looking for any meaningful connection. The data that algorithms train on as well as the predictions or recommendations they output are predetermined.
- **Semi-supervised learning:** This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labeled training data, but the model is free to explore the data on its own and develop its own understanding of the data set.
- **Reinforcement learning:** Data scientists typically use reinforcement learning to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

1.3.1 Main Parts of Machine Learning

- **A Decision Process:** In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labelled or unlabeled, your algorithm will produce an estimate about a pattern in the data.
- **An Error Function:** An error function serves to evaluate the prediction of the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.
- **An Model Optimization Process**: If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model estimate. The algorithm will repeat this evaluate and optimize process, updating weights autonomously until a threshold of accuracy has been met.

1.4 Skin Disease

- Skin disorders are illnesses that affect the surface of your skin. Rashes, inflammation, itching, and other skin abnormalities may be caused by several disorders. Some skin disorders are caused by genetics, while others are caused by lifestyle factors. Medications, lotions, and ointments, as well as lifestyle changes, may be used to treat skin diseases. The big organ that covers and protects your body is your skin. Your skin serves a variety of purposes. It accomplishes the following:
- Prevent dehydration by retaining fluid.
- Assist you in sensing feelings such as temperature and discomfort.
- Prevent the spread of germs, viruses, and other infectious agents.

- Maintain a constant body temperature.
- In reaction to sun exposure, synthesise (produce) vitamin D.
- All problems that clog, irritate, or inflame your skin are considered skin illnesses. Rashes or other changes in your skin's appearance are common symptoms of skin illnesses.

1.4.1 Common Types of Skin Disease

Some skin diseases are minor. Others cause severe symptoms. Some of the most common skin diseases include:

- Acne, which causes oil, germs, and dead skin to build up in your pores due to blocked skin follicles.
- Alopecia areata (hair loss in tiny spots).
- Eczema (atopic dermatitis), a dry, itchy skin condition that causes swelling, cracking, and scaliness.
- Psoriasis is a scaly skin condition that can enlarge and feel hot.
- Raynaud's phenomenon, which causes numbness or skin colour changes by reducing blood flow to your fingers, toes, or other body parts on a regular basis.
- Rosacea, which causes flushed, thick skin and pimples on the face.
- Skin cancer is characterised by the uncontrolled proliferation of abnormal skin cells.
- Vitiligo is a condition in which regions of skin lose their colour.

1.4.2 Causes of Skin Disease

A skin illness can be caused by a variety of lifestyle factors. Your skin may be affected by underlying health issues as well. Bacteria trapped in your pores or hair follicles are common causes of skin problems.

- Thyroid, kidney, or immune system disorders.
- Exposure to environmental stimuli, such as allergens or the skin of another person.
- Genetics
- The presence of fungus or parasites on your skin.
- Anti-inflammatory drugs, such as those used to treat inflammatory bowel disease (IBD).
- Viruses.
- Diabetes and Sun.

1.4.3 Symptoms of Skin Diseases:

Depending on the type of skin illness you have, the symptoms can be rather different. Changes in the skin aren't usually caused by skin illnesses. Wearing ill-fitting shoes, for example, can cause a blister. Skin changes that appear without a known cause, on the other hand, may be linked to an underlying ailment.

Skin illnesses can result in the following symptoms:

- Discolored skin patches (abnormal pigmentation).
- Skin that is dry.
- Sores, lesions, or ulcers that are open.
- Skin that is peeling.
- Rashes, possibly accompanied by itching or pain.
- Bumps that are red, white, or pus-filled.
- Skin that is scaly or rough.

II. LITERATURE REVIEW

[1] AUTHOR NAME: L.Wei, Q. Gan, and T. Ji, (2018)

TITLE: Skin Disease Recognition Method Based on Image Color and Texture Features.

DESCRIPTION: In this paper, three type skin diseases such as herpes, dermatitis, and psoriasis skin disease could be identified by a new recognition method. Initially, skin images were preprocessed to remove noise and irrelevant background by filtering and transformation. Then the method of grey-level co-occurrence matrix (GLCM) was introduced to segment images of skin disease. The texture and color features of different skin disease images could be obtained accurately. Finally, by using the support vector machine (SVM) classification method, three types of skin diseases were identified. The experimental results demonstrate the effectiveness and feasibility of the proposed method.

[2] AUTHOR NAME: R. Sumithra, S. Mahamad , D. Guru

TITLE: Segmentation and Classification of Skin Lesions for Disease Diagnosis,"

DESCRIPTION: In this paper, a novel approach for automatic segmentation and classification of skin lesions is proposed. Initially, skin images are filtered to remove unwanted hairs and noise and then the segmentation process is carried out to extract lesion areas. For segmentation, a region growing method is applied by automatic initialization of seed points. The segmentation performance is measured with different well known measures and the results are appreciable. Subsequently, the extracted lesion areas are represented by color and texture features. SVM and k-NN classifiers are used along with their fusion for the classification using the extracted features. The performance of the system is tested on our own dataset of 726 samples from 141 images consisting of 5 different classes of diseases. The results are very promising with 46.71% and 34% of F-measure using SVM and k-NN classifier respectively and with 61% of F-measure for fusion of SVM and k-NN.

[3] AUTHOR NAME: S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia,

TITLE: Human skin detection using RGB, HSV and YCbCr Color models

DESCRIPTION: Skin color is often used in human skin detection because it is invariant to orientation and size and is fast to process. A new human skin detection algorithm is proposed in this paper. The three main parameters for recognizing a skin pixel are RGB (Red, Green, Blue), HSV (Hue, Saturation, Value) and YCbCr (Luminance, Chrominance) color models. The objective of proposed algorithm is to improve the recognition of skin pixels in given images. The algorithm not only considers individual ranges of the three color parameters but also takes into ac- count combinational ranges which provide greater accuracy in recognizing the skin area in a given image.

[4] AUTHOR NAME: S. Kalaiarasi , H. Kumar and S. Patra,

TITLE: Dermatological Disease Detection using Image Processing and Neural Networks

DESCRIPTION: In this research paper, we provide an approach to detect various kinds of these diseases. We use a dual stage approach which effectively combines Computer Vision and Machine Learning on clinically evaluated histopathological attributes to accurately identify the disease. In the first stage, the image of the skin disease is subject to various kinds of pre-processing techniques followed by feature extraction. The second stage involves the use of Machine learning algorithms to identify diseases based on the histopathological attributes observed on analysing of the skin. Upon training and testing for the six diseases, the system produced an accuracy of up to 95 percent.

[5] AUTHOR NAME: R. Parikh and D. Shah,

TITLE: A Survey on Computer Vision Based Diagnosis for Skin Lesion Detection

DESCRIPTION: We propose an automated image based system for recognition of skin diseases using machine learning classification. This system will utilize computational technique to analyze, process, and relegate the image data predicated on various features of the images. Skin images are filtered to remove unwanted noise and also process it for enhancement of the image. Feature extraction using complex techniques such as Convolutional Neural Network (CNN), classify the image based on the algorithm of softmax classifier and obtain the diagnosis report as an output. This system will give more accuracy and will generate results faster than the traditional method, making this application an efficient and dependable system for

dermatological disease detection. Furthermore, this can also be used as a reliable real time teaching tool for medical students in the dermatology stream.

Problem Definition

- Rule based approach is used for classification which gives static range value for different classes. Therefore we will not able dynamic images or outlier behavior images.
- Features set is not normalize. Therefore different features show different outputs and show different representation during training phase of classifiers.
- Classifiers are not able to distinguish feature overlapping. Therefore at learning phase pattern of image is not identified.
- Multi classification problem is not optimized and ignore the class imbalance problem. Therefore in learning all classes is not input in learning phase so it became a biased learning.
- In the existing method Ada-boost classifier is being used for classification and PCA is used for image segmentation.

Drawbacks

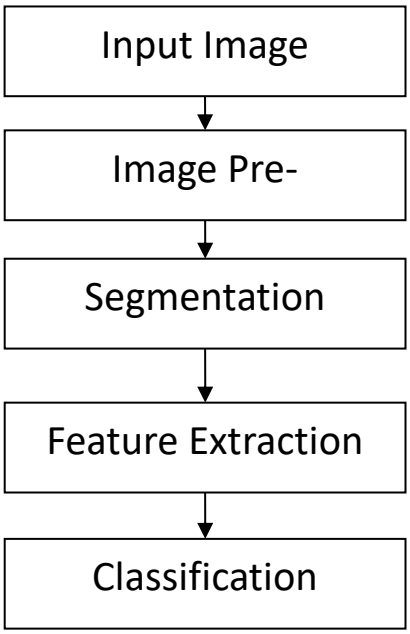
- Poor illumination condition.
- Noisy texture background.
- Inaccurate
- Segmentation not in proper way
- Low performance
- Inefficiency

III. PROPOSED METHOD

- In our proposed system, to classify the skin disease based on the image processing techniques.
- Here, the five steps will be implemented to show the results. They Steps are Datasets collection, Preprocessing, Segmentation, Feature Extraction and Classification.
- Datasets collection can be obtained in the normal and abnormal image format of the system.
- In Preprocessing, the Gaussian filter method is used to show the results.
- In Segmentation, the FCM and Edge detection method is used. In the Feature Extraction part, the DWT and GLCM method is used to show the results.
- In the Classification part, the SVM techniques are used to classify the skin disease based on the Image Processing techniques.
- Finally, the performance metrics have been obtained to show the results.

Advantages

- Better Performance.
- Accuracy is more
- Precision is more



Modules Description

1. **Input Image:** A Collection of data is called datasets. Let us consider, Skin disease databases namely, normal and abnormal prediction of the patient. Here, an input data can be obtained in the image format to predict the disease based on the Skin Disease Detection. 2 Classes of datas will be taken based on the Normal and abnormal condition.
2. **Preprocessing:** Data cleaning, smoothing, grouping or Filtering the image. Data can require preprocessing techniques to ensure accurate, efficient, or meaningful analysis. Data cleaning refers to methods for finding, removing, and replacing bad or missing data. Here, the Gaussian method is used in the Preprocessing techniques. A Gaussian filter is used to reduce the noise in the input image of the system.

2.1 Gaussian Filter

In electronics and signal processing, a Gaussian filter is a filter whose impulse response is a Gaussian function (or an approximation to it, since a true Gaussian response is physically unrealizable as it has infinite support). Gaussian filters have the properties of having no overshoot to a step function input while minimizing the rise and fall time. This behavior is closely connected to the fact that the Gaussian filter has the minimum possible group delay. It is considered the ideal time domain filter, just as the sinc is the ideal frequency domain filter. These properties are important in areas such as oscilloscopes and digital telecommunication systems.

A Gaussian filter is used to reduce the noise in the input image of the system. The Gaussian function is an infinite window length. However, since it decays rapidly, it is often reasonable to truncate the filter window and implement the filter directly for narrow windows, in effect by using a simple rectangular window function. In other cases, the truncation may introduce significant errors. Better results can be achieved by instead using a different window function.

IV. SEGMENTATION

Segmentation means partitions a picture with comparable characteristics into different areas comprising each pixel. The areas should be heavily related to depicted objects or characteristics of interest in order to be relevant and helpful for picture assessment and interpretation. Significant segmentation is the first step from low-level picture processing to transform a gray or color picture into one or more other pictures into a high-level picture description of characteristics, items, and scenes. Successful image analysis relies on segmentation reliability, but precise picture partitioning is usually a very difficult issue. Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). The goal of segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in

images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. Here, segmentation can be used in the FCM of the system.

4.1 FCM

A FCM is a one type of technique is used to segment the image from the input data. Fuzzy clustering also referred to as soft clustering or soft *k*-means is a form of clustering in which each data point can belong to more than one cluster. Clustering or cluster analysis involves assigning data points to clusters such that items in the same cluster are as similar as possible, while items belonging to different clusters are as dissimilar as possible. Clusters are identified via similarity measures. These similarity measures include distance, connectivity, and intensity. Different similarity measures may be chosen based on the data or the application.

V. FEATURE EXTRACTION

Feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images presented as pixels), then it can be transformed into a reduced set of features is also named a feature vector. Determining a subset of the initial features is called feature selection. The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data.

Feature extraction involves reducing the number of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power, also it may cause a classification algorithm to overfit to training samples and generalize poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Here, Feature Extraction method can be obtained in the DWT and GLCM.

5.1 DWT

In numerical analysis and functional analysis, a discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution and it captures both frequency and location information (location in time).

5.2 GLCM

A co-occurrence matrix or co-occurrence distribution is also referred to as gray-level co-occurrence matrices GLCMs) is a matrix that is defined over an image to be the distribution of co-occurring pixel values (grayscale values, or colors) at a given offset. It is used as an approach to texture analysis with various applications especially in medical image analysis.

5.3 Classification

Classification is a term used both about the process to classify disease based on the normal and abnormal of the patient. A classification process can be obtained to training the network of the system. Here, a classification technique can be obtained from using SVM model to classify the result of the model. Finally, the performance of the system can be analysed.

VI. APPLICATION

- Image sharpening and restoration
- Medical field
- Remote sensing
- Transmission and encoding
- Machine/Robot vision
- Color processing

- Pattern recognition
- Video processing
- Microscopic Imaging

6.1 Image sharpening and restoration

Image sharpening and restoration refers here to process images that have been captured from the modern camera to make them a better image or to manipulate those images in way to achieve desired result. It refers to do what Photoshop usually does. This includes Zooming, blurring , sharpening , gray scale to color conversion, detecting edges and vice versa , Image retrieval and Image recognition.

6.2 Medical field

The common applications of DIP in the field of medical is

- Gamma ray imaging
- PET scan
- X Ray Imaging
- Medical CT
- UV imaging

6.3 UV imaging

In the field of remote sensing , the area of the earth is scanned by a satellite or from a very high ground and then it is analyzed to obtain information about it. One particular application of digital image processing in the field of remote sensing is to detect infrastructure damages caused by an earthquake. As it takes longer time to grasp damage, even if serious damages are focused on. Since the area effected by the earthquake is sometimes so wide , that it not possible to examine it with human eye in order to estimate damages. Even if it is , then it is very hectic and time-consuming procedure. So, a solution to this is found in digital image processing. An image of the affected area is captured from the above ground and then it is analyzed to detect the various types of damage done by the earthquake.

VII. CONCLUSION

In this paper, we performed the image processing four segmentation techniques to classify the skin disease to be informative regarding the detailed information relative to the images. The proposed method implemented the Machine learning to show the results. Here, the abnormal condition based on the different disease images, they are Acne and Melonama and Normal skin image. The input image is considered to be a set of data. The resultant set is implemented and structured by using Matlab Software. It will help us to find the most suitable classification technique of the following to detect the skin diseases. From our experimental result, SVM model is best algorithm to classify the skin infection in the body. Experimental results can be obtained in the performance metrics of our system.

REFERENCES

[1] Zhang J, Han J, Liu T, Zhou J, Suni and Luo J. Safety prediction of rail transit system based on deep learning, 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan. 2017;851-856.

[2] Anitha G.K, Deepak M.C. Machine Learning Techniques for learning features of any kind of data: A Case Study. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET).* 2014 December;3(12).

[3] Yasir, R., Rahman, A., & Ahmed, N. (2014) "Dermatological Disease Detection using Image Processing and Artificial Neural Network. "Dhaka: International Conference on Electrical and Computer Engineering.

[4] Suganya R. (2016) "An Automated Computer Aided Diagnosis of Skin Lesions Detection and Classification for Dermoscopy Images." International Conference on Recent Trends in Information Technology.