DERMATALOGICAL MANIFESTATIONS SEGMENTATION USING MACHINE LEARNING

Submitted in partial fulfillment of the
requirements for the award of
Bachelor of Engineering degree in Computer Science and Engineering

By

DANDA SRI VAMSI (Reg.No – 39110243)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHISALAI,
CHENNAI - 600119

NOVEMBER - 2022

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with —A‖ grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **DANDA SRI VAMSI (39110243)** who carried out the Project Phase-1 entitled **"DERMATOLOGICAL MANIFESTATIONS SEGMENTATION USING MACHINE LEARNING"** under my supervision from June 2022 to November 2022.

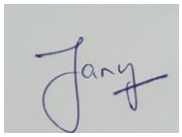
Internal Guide

Dr. JOSHILA GRACE L.K

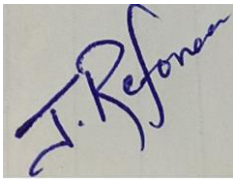
Head of the Department

Dr. L. LAKSHMANAN, M.E., Ph.D.

Submitted for Viva voce Examination held on 28-04-2023



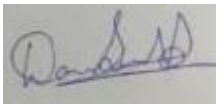
InternalExaminer



ExternalExaminer

DECLARATION

I, **DANDA SRI VAMSI (39110243)** hereby declare that the Project Phase-1 Report entitled **"DERMATOLOGICAL MANIFESTATIONS SEGMENTATION USING MACHINE LEARNING"** done by me under the guidance of **Dr.JOSHILA GRACE L.K** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Computer Science and Engineering**.



DATE: 28-10-2022
PLACE: Chennai **SIGNATURE OF THECANDIDATE**

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph. D**, **Dean**, School of Computing, **Dr. L. Lakshmanan M.E., Ph. D.,** Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.JOSHILA GRACE.L.K** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my phase-1 project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

In recent days, skin diseases have been significantly increasing in humans. Skin diseases are manifestated either as infections or allergies. Some of the skin diseases such as ring worm infections, yeast infections, viral infections, eczemas etc. are spreading rapidly day by day. These type of diseases need to be identified in their earlier stages to avoid spreading. Some factors like clinical parameters are considered for identifying the disease. The possible skin diseases in different ages are dermatitis in age 0-5 years, warts affects in 6-11 years age, and acne vulgaris in 12-16 years age. Dermatomyositis is a type of skin disease that affects children at age of 5-15 and adults at 40-60 age.

.

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

There are different techniques to diagnose skin disease like Malignant melanoma i.e. melanoma skin tumors, Nonmelanoma i.e. squamous cell carcinoma, basal cell carcinoma, Acne, Genetic Diseases i.e. Sickle-Cell Anemia Genetic Skin Disorder, Fungal Diseases, Bacterial Diseases, Psoriasis and Leprosy. There are different techniques to diagnose these skin diseases. But the most useful and noninvasive method for diagnosis of different types of skin diseases is skin impedance measurement. The analysis is made with the measurement of skin impedance over a wide frequency range from 100 Hz to 1MHz. But this analysis for the infected skin and normal skin should be made at same frequency because at different frequencies skin impedance can be different. The measured skin impedance at same frequency of infected skin and normal skin is different, because of which we can diagnose skin disease. Above mentioned method used the non-invasive probe electrode system consists of two concentric electrodes attached to a ceramic plate. The outermost electrodes drive the voltage and inner is sink electrode gives output. In which source electrode provide small amount of current to skin and sink electrode gives input to the microcontroller. Length of the outermost electrode is nearly 10mm. Measurement of skin impedance can be used for depressions, anxiety, neurological diseases, and nerve lesions. It is also used for physiological measurements emotional disorders or lie detection.

The various types of bioimpedance method include multiple frequency, single frequency and Bio impedance spectroscopy (SF-BIA). SF-BIA is generally performed at a frequency of 50 kHz. At this frequency the current passes through both extracellular and intracellular fluid. Bio-impedance obtained at number of frequencies is analyzed in multiple-frequency bio-impedance analysis.

1.1. DOMAIN OVERVIEW :

Machine Learning is the most popular technique of predicting the future or classifying information to help people in making necessary decisions. Machine Learning algorithms are trained over instances or examples through which they learn from past experiences and also analyze the historical data. Therefore, as it trains over the examples, again and again, it is able to identify patterns in order to make predictions about the future. Data is the core backbone of machine learning algorithms. With the help of the historical data, we are able to create more data by training these machine learning algorithms. For example, Generative Adversarial Networks are an advanced concept of Machine Learning that learns from the historical images through which they are capable of generating more images. This is also applied towards speech and text synthesis. Therefore, Machine Learning has opened up a vast potential for data science applications.

1.2 MACHINE LEARNING

Machine Learning combines computer science, mathematics, and statistics. Statistics is essential for drawing inferences from the data. Mathematics is useful for developing machine learning models and finally, computer science is used for implementing algorithms. However, simply building models is not enough. You must

also optimize and tune the model appropriately so that it provides you with accurate results. Optimization techniques involve tuning the hyper parameters to reach an optimum result.

The world today is evolving and so are the needs and requirements of people. Furthermore, we are witnessing a fourth industrial revolution of data. In order to derive meaningful insights from this data and learn from the way in which people and the system interface with the data, we need computational algorithms that can churn the data and provide us with results that would benefit us in various ways. Machine Learning has revolutionized industries like medicine, healthcare, manufacturing, banking, and several other industries. Therefore, Machine Learning has become an essential part of modern industry. Data is expanding exponentially and in order to harness the power of this data, added by the massive increase in computation power, Machine Learning has added another dimension to the way we perceive information. Machine Learning is being utilized everywhere. The electronic devices you use, the applications that are part of your everyday life are powered by powerful machine learning algorithms.

With an exponential increase in data, there is a need for having a system that can handle this massive load of data. Machine Learning models like Deep Learning allow the vast majority of data to be handled with an accurate generation of predictions. Machine Learning has revolutionized the way we perceive information and the various insights we can gain out of it. These machine learning algorithms use the patterns

contained in the training data to perform classification and future predictions. Whenever any new input is introduced to the ML model, it applies its learned patterns over the new data to make future predictions. Based on the final accuracy, one can optimize their models using various standardized approaches. In this way, Machine Learning model learns to adapt to new examples and produce better results.

Types of Machine Learning

Machine Learning Algorithms can be classified into 3 types as follows –

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

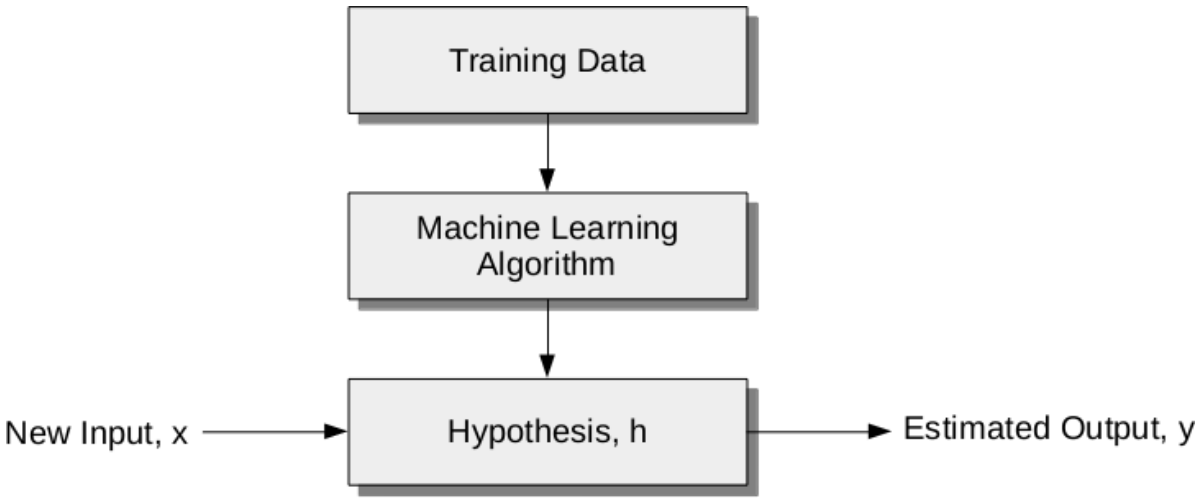


Fig-1 Supervised Architecture

1.3 SUPERVISED LEARNING

In the majority of supervised learning applications, the ultimate goal is to develop a finely tuned predictor function h(x) (sometimes called the “hypothesis”). “Learning” consists of using sophisticated mathematical algorithms to optimize this function so that,

given input data x about a certain domain (say, square footage of a house), it will accurately predict some interesting value h(x) (say, market price for said house).

$$h(x_1, x_2, x_3, x_4) = \theta_0 + \theta_1 x_1 + \theta_2 x_3^2 + \theta_3 x_3 x_4 + \theta_4 x_1^3 x_2^2 + \theta_5 x_2 x_3^4 x_4^2$$

This function takes input in four dimensions and has a variety of polynomial terms. Deriving a normal equation for this function is a significant challenge. Many modern machine learning problems take thousands or even millions of dimensions of data to build predictions using hundreds of coefficients. Predicting how an organism's genome will be expressed, or what the climate will be like in fifty years, are examples of such complex problems.

Under supervised ML, two major subcategories are:

- Regression machine learning systems: Systems where the value being predicted falls somewhere on a continuous spectrum.
- Classification machine learning systems: Systems where we seek a yes-or-no prediction.

In practice, x almost always represents multiple data points. So, for example, a housing price predictor might take not only square-footage (x1) but also number of bedrooms (x2), number of bathrooms (x3), number of floors (x4), year built (x5), zip code (x6), and so forth. Determining which inputs to use is an important part of ML design. However, for the sake of explanation, it is easiest to assume a single input value is used

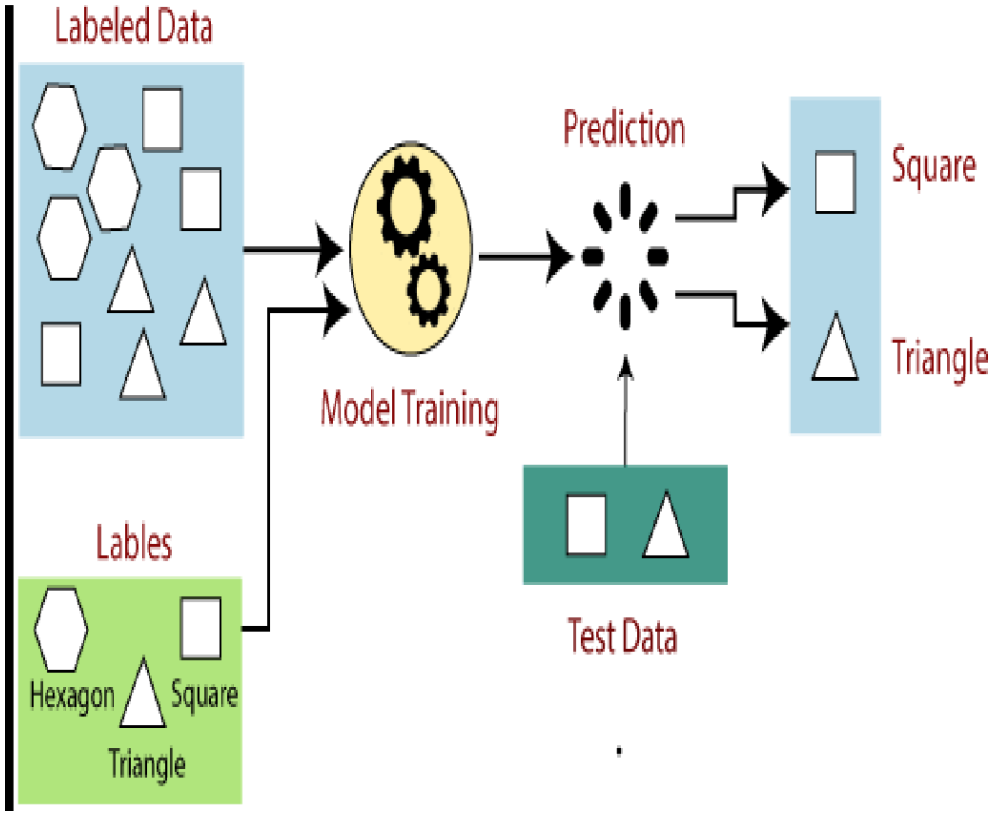


Fig-2 Block Diagram

Steps Involved in Supervised Learning:

- First Determine the type of training dataset
- Collect/Gather the labelled training data.
- Split the training dataset into training dataset, test dataset, and validation dataset.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.

- Evaluate the accuracy of the model by providing the test set.

1.4 REGRESSION

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc.

1. Linear Regression
2. Regression Trees
3. Non-Linear Regression
4. Bayesian Linear Regression
5. Polynomial Regression

CHAPTER 2

LITERATURE SURVEY

2.1. Segmentation and Classification of Skin Lesions for Disease Diagnosis

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.

2.2. Extraction of Skin Lesions from Non-Dermoscopic Images Using Deep Learning

Melanoma is amongst most aggressive types of cancer. However, it is highly curable if detected in its early stages. Prescreening of suspicious moles and lesions for malignancy is of great importance. Detection can be done by images captured by standard cameras, which are more preferable due to low cost and availability. One important step in computerized evaluation of skin lesions is accurate detection of lesion region, i.e. segmentation of an image into two regions as lesion and normal skin. Accurate segmentation can be challenging due to burdens such as illumination variation and low contrast between lesion and healthy skin. In this paper, a method based on deep neural networks is proposed for accurate extraction of a lesion region. The input image is preprocessed and then its patches are fed to a convolutional neural network. Local texture and global structure of the patches are processed in order to assign pixels to lesion or normal classes. A method for effective selection of training patches is used for more accurate detection of a lesion border. The output segmentation mask is refined by some post processing operations. The experimental results of qualitative and quantitative evaluations demonstrate that our method can outperform other state-of-the- art algorithms exist in the literature.

2.3 An intelligent system to diagnosis the skin disease

Skin cancer is a major health issue affecting a vast segment of the population regardless the skin color. This affectation can be detected using dermoscopy to determine whether the visible spots on skin are either benign or malignant tumors. In spite of the specialists' experience, skin lesions are difficult to classify, reason for which

computer systems are developed to increase the effectiveness of cancer detection. Systems assisting in the detection of skin cancer process digital images to determine the occurrence of tumors by interpreting clinical parameters, relying, firstly, upon an accurate segmentation process to extract relevant features. Two of the well-known methods to analyze lesions are ABCD and the 7-point check list. After clinically-relevant features are extracted, they are used to classify the presence or absence of a tumor. However, irregular and disperse lesion borders, low contrast, artifacts in images and the presence of various colors within the region of interest complicate the processing of images. In this article, we propose an intelligent system running the following method. The feature extraction stage begins with the segmentation of an image, for which we apply the Wavelet – Fuzzy C-Means algorithm. Next, specific features should be determined, among others the area and the asymmetry of the lesion. An ensemble of clusterers extracts the Red-Green-Blue values that correspond to one or more of the colors defined in the ABCD guide. The feature extraction stage includes the discovery of structures that appear in the lesion according to the method known as Grey Level Cooccurrence Matrix. Then, during the detection phase, an ensemble of classifiers determines the occurrence of a malignant tumor. Our experiments are performed on images taken from the ISIC repository. The proposed system provides a skin cancer detection performance above 88 percent, as measured by the accuracy. Details of how this performance fares when compared with other systems are also given.

2.4 Automatic segmentation of skin lesions from dermatological photograph

Melanoma is the deadliest form of skin cancer if left untreated. Incidence rates of melanoma have been increasing, especially among young adults, but survival rates are

high if detected early. Unfortunately, the time and costs required for dermatologists to screen all patients for melanoma are prohibitively expensive. There is a need for an automated system to assess a patient's risk of melanoma using photographs of their skin lesions. Dermatologists could use the system to aid their diagnosis without the need for special or expensive equipment. One challenge in implementing such a system is locating the skin lesion in the digital image. Most existing skin lesion segmentation algorithms are designed for images taken using a special instrument called the dermatoscope. The presence of illumination variation in digital images such as shadows complicates the task of finding the lesion. The goal of this research is to develop a framework to automatically correct and segment the skin lesion from an input photograph. The first part of the research is to model illumination variation using a proposed multi-stage illumination modeling algorithm and then using that model to correct the original photograph. Second, a set of representative texture distributions are learned from the corrected photograph and a texture distinctiveness metric is calculated for each distribution. Finally, a texture-based segmentation algorithm classifies regions in the photograph as normal skin or lesion based on the occurrence of representative texture distributions. The resulting segmentation can be used as an input to separate feature extraction and melanoma classification algorithms. The proposed segmentation framework is tested by comparing lesion segmentation results and melanoma classification results to results using other state-of-the-art algorithms. The proposed framework has better segmentation accuracy compared to all other tested algorithms. The segmentation results produced by the tested algorithms are used to train an existing classification algorithm to identify lesions as melanoma or non-melanoma. Using the proposed framework produces the highest classification accuracy and is tied

for the highest sensitivity and specificity.

2.5 A Survey on skin texture analysis for medical diagnosis using image processing techniques

Skin texture analysis is one of then challenging issues in the field of medical diagnosis. Various types of skin diseases are affecting human life like skin dryness, fungus, and allergic symptoms. The objective of this paper is to analyze the skin disease using texture analysis of skin image and by comparing the test image to a defined images or reference images. The matching of test and reference images compared that yields the percentage of skin diseases in the captured skin texture image.

2.6 Existing system

In this article, existing method that uses computer vision based techniques to detect various kinds of dermatological skin diseases. We have used different types of machine learning algorithms for feature extraction and feed forward artificial neural network for training and testing purpose. The system works on two phases- first pre-process the colour skin images to extract significant features and later identifies the diseases. The system successfully detects 9 different types of dermatological skin diseases with an accuracy rate of 90%.

Disadvantages of existing system

- Does not analyse the disease
- Less security
- There is no feedback system

2.7 Proposed system

In proposed system, KNN and Naïve Bayes techniques to detect various kinds of skin diseases. It can be identified using some factors like clinical parameters which are considered for identifying the disease. The possible skin diseases in different ages are dermatitis in age 0-5 years, warts affects in 6-11 years age, and acne vulgaris in 12-16 years age. Dermatomyositis is a type of skin disease that affects children at age of 5-15 and adult at 40-60 age. We are giving a suggestion of what kind of treatment you want and also showing life thread or not.

Advantages of proposed system:

- Easily analyze the disease
- High Security

CHAPTER 3
SYSTEM REQUIREMENTS

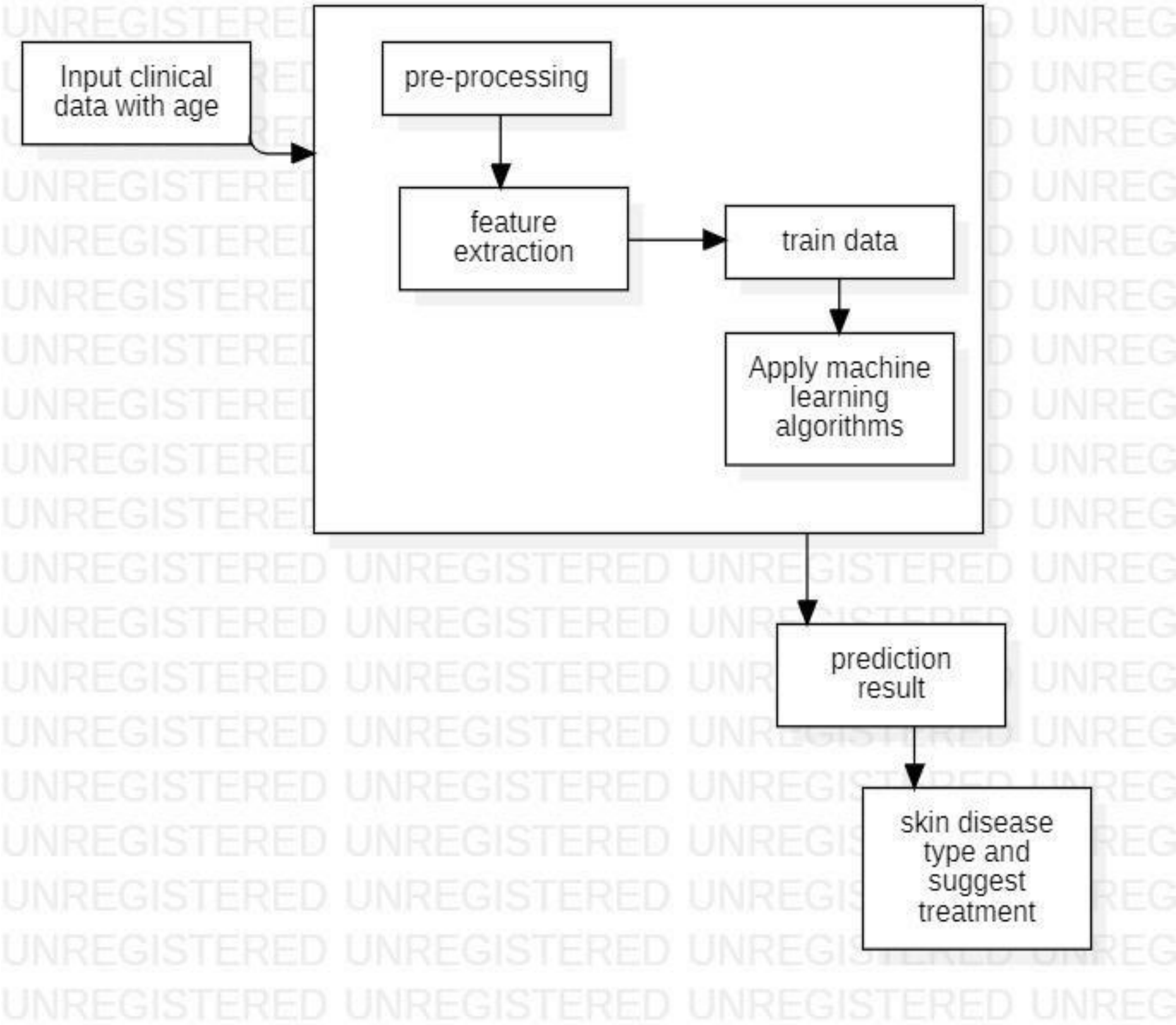
HARDWARE REQUIREMENTS:

- Hard Disk : 100 GB.
- Input Devices: Keyboard, Mouse
- Ram : 4GB.

SOFTWARE REQUIREMENTS:

- Operating system : Windows 10
- Coding Language : Java
- Toolkit : Netbeans
- DATABASE : MySql

SYSTEM ARCHITECTURE



REFERENCES

[1] Sumithra Ra, Mahamad Suhilb, D.S.Guruc," Segmentation and Classification of Skin Lesions for Disease Diagnosis", doi: 10.1016/j.procs.2015.03.090, 2015.

[2] Mohammad H. Jafari, Ebrahim Nasr-Esfahani, Nader Karimi, S.M. Reza Soroushmehr, Shadrokh Samavi, Kayvan Najarian, " Extraction of Skin Lesions from Non-Dermoscopic Images Using Deep Learning", ResearchGate, September 2016.

[3] Manish Kumar and Rajiv Kumar, "An intelligent system to diagnosis the skin disease", vol. 11, no. 19, october 2016.

[4] Jerrey Glaister," Automatic segmentation of skin lesions from dermatological photograph", Waterloo, Ontario, Canada, 2013.

[5] K.indupriya, Dr. G.P.Ramesh kumar, "A survey on skin texture analysis for medical diagnosis using image processing techniques", K.Indupriya et al. Volume 3 Issue 5, 2015.

[6] Md Nafiul Alam, Tamanna Tabassum Khan Munia, Kouhyar Tavakolian, Vasefi, Nick MacKinnon, Reza Fazel-Rezai, "Automatic Detection and Severity Measurement of Eczema Using Image Processing", IEEE, 2016.