A project report on

AN INTELLIGENT SYSTEM FOR HAIR IMAGE SCALP DISEASE DETECTION AND DIAGNOSIS

Submitted in partial fulfillment for the award of the degree of

Bachelor of Technology

by

VARSHA R(18BCE1344)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

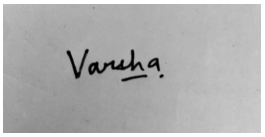
MAY , 2022

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

DECLARATION

I hereby declare that the thesis entitled "AN INTELLIGENT SYSTEM FOR HAIR IMAGE SCALP DISEASE DETECTION AND DIAGNOSIS " submitted by me, for the award of the degree of Bachelor of Technology in Computer Science and Engineering, Vellore Institute of Technology is a record of bonafide work carried out by me under the supervision of Dr. Asha S.

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.



Place: Chennai

Date: 03/05/2022

Signature of the Candidate

VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

CERTIFICATE

This is to certify that the report entitled **"AN INTELLIGENT SYSTEM FOR HAIR IMAGE SCALP DISEASE DETECTION AND DIAGNOSIS"** is prepared and submitted by **VARSHA R(18BCE1344)** to VIT Chennai, in partial fulfillment of the requirement for the award of the degree of **B.Tech. CSE** programme is a bonafide record carried out under my guidance. The project fulfills the requirements as per the regulations of this University and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma and the same is certified.

Signature of the Guide:

Name: Dr. Asha S

Date:

Signature of the Internal Examiner

Name:

Date:

Signature of the External Examiner

Name:

Date:

Approved by the Head of Department, **B. Tech CSE**

Name: Dr. Nithyanandam P

Date:

(Seal of SCOPE)

ABSTRACT

In recent times, hair care has become exceedingly exorbitant. Since breakthroughs in intelligent device computation capability as well as reduced prices, it has become achievable to get a low-cost hair follicles monitoring system.

Baldness has created a substantial social issue for people and has created a drastic impact on peoples lives.Thus, hair scalp analysis is an efficient predictive approach for monitoring the health of internal organs which can be used for detecting illness associated with the scalp. Hair care analysis done at a professional hair care center is highly expensive. Keeping this in mind, we have devised an intelligent system to track illnesses on the go. Multiple features including textural, geometric, color and hair count have been extracted from the scalp images, which were further utilized to train Machine Learning algorithms to identify Melanoma, Alopecia Areata and Tinea Capitis. This main focus of this research will automatically classify the condition of the patient's hairline. Moreover, through implementing machine learning algorithms, we can progressively expand the number of observations in order to promote efficiency.

Hair texture, quantity, and scalp patch are some of the attributes. Rigorous testing upon that prototype design assessed the effectiveness of our concept. As an outcome, we have collected quantitative information about scalp, such as bacteria, hypersensitivity, flaking, greasing, and baldness.
Medical professionals would be able to attain independent advice using the proposed approach, which also will help them in making accurate assessments for identifying the prevalence of this disease in people.

ACKNOWLEDGEMENT

It is my pleasure to express with deep sense of gratitude to Dr. Asha S,
Associate Professor, SCOPE, Vellore Institute of Technology, Chennai, for her constant guidance, continual encouragement, understanding; more than all, she taught me patience in my endeavor. My association with her is not confined to academics only, but it is a great opportunity on my part to work with an intellectual and expert in the field of Machine Learning.

It is with gratitude that I would like to extend thanks to our honorable Chancellor Dr. G. Viswanathan, all the vice president's Mr. Sankar Viswanathan, Dr. Sekar Viswanathan and Mr. G V Selvam, Assistant Vice-President Ms. Kadhambari S. Viswanathan, Vice-Chancellor Dr. Rambabu Kodali and Pro-Vice Chancellor Dr. V. S. Kanchana Bhaaskaran for providing an exceptional working environment and inspiring all of us during the tenure of the course.

In a jubilant mood I express ingeniously my whole-hearted thanks to Dr. Nithyanandam P, Head of the Department, Dr. Karmel A, Co Chair, and Project Coordinator Dr. Abdul Quadir Md, B. Tech. Computer Science and Engineering, SCOPE, Vellore Institute of Technology, Chennai for their valuable support and encouragement to take up and complete the thesis.

Special mention to Dean, Dr. Ganesan R, Associate Dean, Dr. Geetha S, SCOPE, Vellore Institute of Technology, Chennai for spending their valuable time and efforts in sharing their knowledge and for helping us in every aspect.

All teaching staff and members working as members of our university prompted the acquisition of the requisite knowledge to finalize my course study successfully. I would like to thank my parents for their support.

It is indeed a pleasure to thank my friends who encouraged me to take up and complete this task. At last but not least, I express my gratitude to all those who have helped me directly or indirectly toward the successful completion of this project.

Place: Chennai

Date: May, 2022 **VARSHA R**

TABLE OF CONTENTS

CONTENTS iv

LIST OF FIGURES ix

LIST OF TABLES xi

LIST OF ACRONYMS xii

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION 10

1.2 PROJECT OVERVIEW 10

1.3 DISEASE CLASSIFICATION 11

1.4 PROJECT GOALS 12

1.5 PROJECT CHALLENGES 12

CHAPTER 2

LITERATURE

SURVEY

2.1 RESEARCH OVERVIEW 15

2.2 EXISTING TECHNOLOGICAL SOLUTIONS 16

2.3 TECHNOLOGICAL DECISIONS 19

2.4 DATASET DESCRIPTION 20

CHAPTER 3

INTRODUCTION

3.1 OVERVIEW 23

3.2 ARCHITECTURE 24

3.3 MATHEMATICAL EQUATION 25

3.4 FINAL REMARKS 28

CHAPTER 4

METHODOLOGY

AND APPROACH

4.1 OVERVIEW 29

4.2 AGILE METHOD 29

4.3 CRISP DM 31

4.4 MODELING 32

CHAPTER 5

SYSTEM DESIGN

5.1 OVERVIEW 34

5.2 USE CASE DIAGRAM 35

5.3 CLASS DIAGRAM 37

5.4 SEQUENCE DIAGRAM 38

5.5 ACTIVITY DIAGRAM 39

CHAPTER 6

IMPLEMENTATI

ON

6 OVERVIEW 40

6.1 SCALP DISEASE DETECTION 42

6.2 IMAGE PROCESSING 44

6.3 IMAGE SEGMENTATION 45

6.4 FEATURE EXTRACTION 46

6.5 SKELETONISATION 48

6.6 HAIR COUNT 49

6.7 TEXTURAL FEATURES 50

CHAPTER 7

CODE AND RESULTS

7.1 DATASET 51

7.2 FEATURES 53

7.3 MACHINE LEARNING MODELS 61

CHAPTER 8

PROJECT PLAN

8.1 INTRODUCTION 69

8.2 GANTT CHART PLAN 69

8.3 PROJECT PLAN OVERVIEW 70

8.4 CONCLUSION 70

CHAPTER 9

CONCLUSION

9.1 INTRODUCTION 71

9.2 ALGORITHMS USED 71

9.3 MODEL RESULTS 72

REFERENCES 73

LIST OF FIGURES

Figure 1: Reasons of hair loss 16
Figure 2 : Alopecia Areata 18
Figure 3 :Tinea capitis 20
Figure 4 :Alopecia Areata 20
Figure 5 :Melanoma 20
Figure 6 : Healthy scalp 21
Figure 7 : ML Model 23
Figure 8 : SVM 24
Figure 9: Optimal Hyperplane using the SVM algorithm 25
Figure 10 : LR ` 26
Figure 11: LR Architecture 27
Figure 12 : LR math model 27
Figure 13 : KNN model 27
Figure 14 : KNN Architecture 28
Figure 15 : CRISP DM 31
Figure 16 : Use case diagram for overall model 35
Figure 17 : Use case diagram for Image processing 36
Figure 18 : Class diagram 37
Figure 19 : Sequence diagram 38
Figure 20 : Activity diagram 39
Figure 21 : Flowchart 41
Figure 22 : Processes of image processing 42
Figure 23 :Feature extraction 43
Figure 24 :Dataset 44
Figure 25 :Grayscale 44
Figure 26 :Segmented image 45
Figure 27 :Color feature 46
Figure 28 :Skeletonisation 48
Figure 29 :Hair count 49

LIST OF TABLES

Table 1 : Inference for color feature 55
Table 2 : Inference for geometric feature 57
Table 3 : Inference for skeletonisation 58
Table 4 : Inference for hair count 59
Table 5 : Inference for textural feature 61

LIST OF SYMBOLS, ABBREVIATIONS AND NOMENCLATURE

ML Machine Learning

SVM Support Vector Machine

LR Logistic regression

KNN

Chapter 1

INTRODUCTION

1.1 INTRODUCTION

The chapter aims to provide a concise summary of the work, such as its objectives, limitations, and document format.

1.2 PROJECT OVERVIEW

1.2.1 IMAGE PREPROCESSING

Prior to retrieving any relevant features, a monochrome conversion is done. Regarding computationally homogeneity and modeling stability, pictures are scaled to 560x560. The (5) × (5) kernel has been used as a two dimensional filtering for sound insulation to mitigate underlying technological deficiencies of such a record and thus to distinguish the desired physiological responses among interfering actions.

1.2.2 IMAGE SEGMENTATION

Utilizing the GrabCut algorithm, we fragment the images foreground-background retrieval. Due to the obvious non-uniformity of the background, the GrabCut technique is recommended over all other methods. This method determines quite well in a circumstance, which really is related to the error model, that stipulates that now the coordinates and wavelengths of the transmission can indeed be determined at the same moment.

1.2.3 FEATURE EXTRACTION

Scalp botch (going bald) utilizing skeletonisation, hair count utilizing hough transform methodology, texture to recognise coating, surface quality, line segments and small pores, color and spatial features such as region, altitude diameter proportion, core width, pivotal height, relatively small ½ length, circular area and circular pattern area ratio.

1.2.4 DISEASE CLASSIFICATION

These image attributes are used to construct three separate classifier models: Support Vector Machine, K - nearest neighbors, and Logistic Regression.

1.2.5 RESEARCH ELEMENT

This study offers a wide range of results of the study that aid significant Machine learning's capability to identify features from the scalp and classify those into known ailments. The scalp database used it to create and train the classifier was compiled by manually collecting it from general citizens and other accessible datasets. Given today's day and age, when interpersonal communication are restricted and prompt disease detection is critical, it is essential to engage in and produce effective and efficient diagnostic techniques.

Deep learning and image analysis strategies are investigated by several researchers to ensure that now the suitable design modeling would be used to collect information and train a model. Even though the KNN has been the most normally employed design in disease identification, studies have also focused on different learning algorithms like the Logistic Regression and Support Vector Machine. This project aims to assist with analyzing a picture of a scalp and generating judgments, thereby helping the experts in the network without any need for human involvement.

1.3 PROJECT GOALS

The finished work ought to be capable of accomplishing the following objectives:

• Programs in Machine - learning Training and OpenCV Training: To have this program began, I did a full amount of self. Originally, my knowledge of Computer Vision and OpenCV was minor and insignificant. Thus, as a consequence, there was a necessity to educate about the topic via various lectures and courses. Prior continuing and putting out any activities, this was necessary to get a solid grasp on the concepts of Computer Vision.
• Scalp picture preprocessing: The system ought to be able to receive & analyze images and videos which are given to that too. So because Vision methods and Machines Learning Model will rely upon that, the system will be able to acquire and process images containing their packet headers.
• Disease identification and classification The system ought to be capable of recognizing diseases from given pictures and classify these into specific categories. It ought to be capable of distinguishing among different scalp photographs and correctly predict.
• As just a result, we will get quantifiable data about the scalp, such as bacteria, allergens, flakes, oil, and loss of hair.
• Using the proposed model, medical experts will be able to obtain a second opinion that will help people make appropriate decisions for diagnosing the presence of this disease in individuals.

1.4 PROJECT CHALLENGES

That chapter discusses the difficulties encountered with completing this project. In the following sections of a study, every subject was shortly described then closely investigated. The phase is crucial to anybody who chooses to run this project with extrinsic pictures.

1.4.1 SEGMENTATION AND PREPROCESSING OF IMAGES

The program's greatest difficult obstacle would definitely become its tremendous scale. Dealing with deformed scalp images makes the data organization and retrieval challenging. Moreover, the computing gets demanding because of prolonged iterations of image processing and analyzing. Since each scalp does have a visible difference, color, and texture, information management and algorithms efficiency is a difficult task.

1.4.2 COMPUTER VISION CONCEPTS

The OpenCV and Machine Learning course is exciting, but it was unknown prior to starting the project. Finishing this project by crossing out all the goals required extensive research & research on the topic. Pursuing various books and the internet courses just on topic were excellent. Classes by renowned academics helped inside the building of a solid foundation of comprehension on Machine Learning.

1.4.3 MODEL TRAINING

Machine Learning algorithms require a significant amount of learning time before they're even used for tests. Conducting testing & receiving results requires significantly longer than normal AI calculations. That dilemma has the potential to impede work progress. That problem can be addressed through proper planning prior to beginning analyses.

1.5 REPORT STRUCTURE

This investigation section discusses the study on advanced diagnosable disorder methods, data on it's own impact, current technologies , remedies, as well as the project's data. A main goal of this section is to highlight that reducing disease risks as early as possible is an inconceivably complex topic that really is presently actively studied.

This section of Machine Learning dives into the various perceptions key concepts required for models learning and training. The chapter includes research papers related to the project which helped inside the construction of models and the solution of various issues which occurred during the course of a project.

The strategy & method chapter has discussed the development's working process. This describes the various activities which were performed through, and the techniques and methodologies that have been used to perform the project. This also discusses different designs utilized in undertaking. That chapter's aim is to verify that now the work is administered in keeping with standard operating procedures.

The designing section will discuss various planning strategies utilized, but also customer plan artifacts such as avatars, occasions, organization schematic design, including specialized structural extreme work like usage scenario chart and frameworks defined schema.

Following the approach & method mentioned previously, this Deployment section addresses the various processes present in the program. It is built just on concepts that were developed as well as the Machine Learning tests which have been put successfully.

This plan describes a construction scope throughout its full length. It covers all of the major phases within the project's development. This also went into great detail about job estimates.

Its epilogue section describes its acquired findings, evidence of concept appraisal, & finally overall future perspectives of that too.

1.6. FINAL REMARKS

The research combines many computer programming concepts in such an attempt to develop a method to produce positive diagnostic - assist technologies. This project involves data gathering and development work that illustrate its concept. The following chapter has discussed this research’s main work.

Chapter 2

LITERATURE SURVEY

2.1 RESEARCH OVERVIEW

The section offers details about project-related studies that have been conducted. This section entails information through various sources on numerous topics that seem to be critical to recognize this endeavor. It requires understanding of a medical diagnosis system. The chapter also discusses both technology used and evidence obtained during the project's development.

2.2 GENERAL STATISTICS

Per the World Academy in Hairstyle Replacement Treatments 2014 survey , approximately 1.2 billion men inside the U. S. suffer from some form of loss of hair, whereas about 21 million women are affected by loss of hair.

With approximately 2.5 billion people in the Usa, as well as at least 20 percent of the overall population being under the age 30, that means that more than one in every five Americans are suffering baldness, especially men account for a higher percentage of baldness sufferers.

This likelihood for men suffering male pattern baldness rises significantly, increasing aging. Even by the age 35, approximately 40percent all men experience visible male pattern baldness, but by the time of 60, that percentage increases almost 65%.

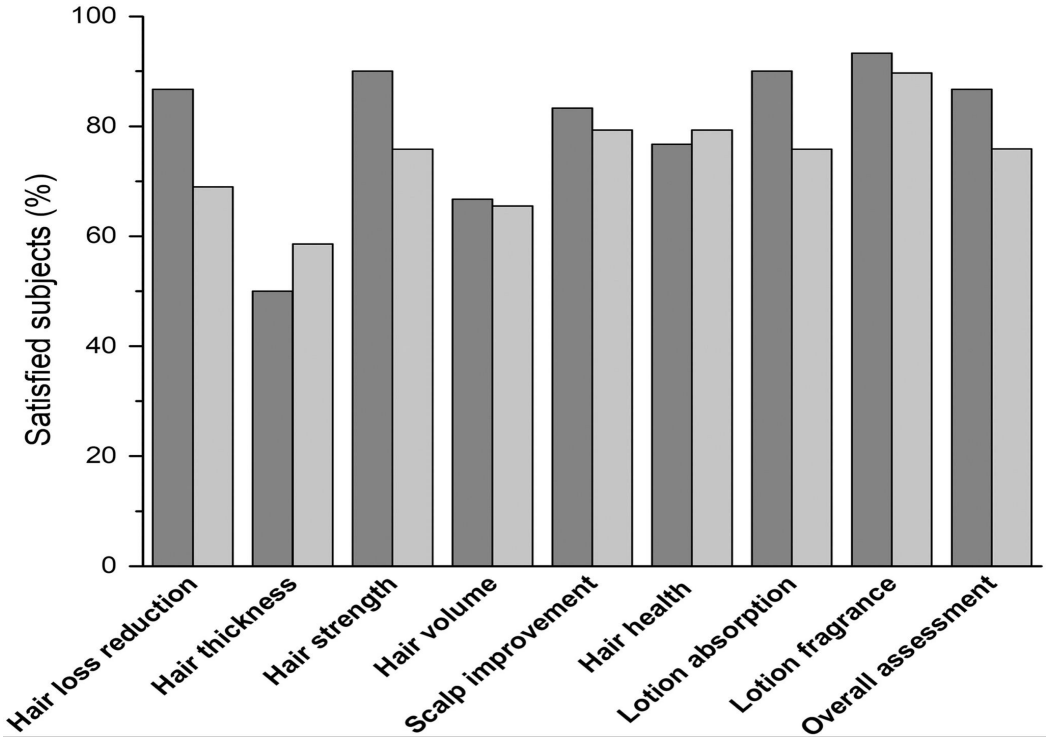


Figure 1: Reasons of hair loss

2.3 EXISTING TECHNOLOGICAL SOLUTIONS

Researchers know that Alopecia is indeed a persistent and recurring condition that causes hair loss. Using CNN architecture, they discovered that degree of baldness is by far the most significant predictor indicator, with Jaccard scores of 0.941 & 0.963 again for hair roots fall region, correspondingly. We will want to use hair growth in the hereafter. [1] .

In the next study, researchers employed a form to poll sport participation black women in order to determine the most common hair styling procedures used to adapt activity and to make recommendations for optimal regimes for health and personal care both during activity. [2] .

This study presents a cell phone hairless forehead diagnosis that employs artificial intelligence such as Ssr or Accelerated R-CNN to accurately recognise & treat flaking, cellulitis, baldness, and unkempt hair. [3] .

This article investigates autonomously recognizing degree of irregular hair from face recognition photographs using data mining algorithms including such Various online sources, Str, and Str Networks and identified 82.5 percent, 84.5 percent, and 85.5 percent accuracy, respectively. [4] .

The goal of the this study is to assess the dermoscopy images features of tinea in kids at higher resolution (150) as well as its diagnosis Those who as well found "Telegraph" hair strands, which do have multiple white rock acts from across hair, and "zig zag pattern" strands of hair, which do have innumerable rock acts at pointed angles2; the above discovery has heretofore been noted in other diseases linked with focus point weakness of a hair, such as alopecia areata as well as trichorrhexis nodosa [5] .

In this research, they present a method for avoiding hair loss and scalp self-diagnosis by extracting HLF (hair loss feature) from a scalp picture captured with a microscope and combining it with grid line selection and eigenvalue. HLF is defined as the quantity of hairs, hair follicles, and hair thickness that include broken hairs, short vellus hairs, and tapering hairs. The number of hairs and hair follicles are counted accurately using a truth label that can be seen with the naked eye. Hair thickness is manually measured with a hand-made hair thickness measurement instrument. The quantity of hair which was not significantly influenced by noise when viewed with the human eye, such as the image itself, is counted. Hairs are represented as a set of itself in the a hair, in save the case of red spots, black polka, and yellow dots which aggravate your tresses from forming. We additionally assessed follicles present on the picture's border but very weakly discernible. [6] .

The deep-learning technique, the Matching models Part - Of - speech (BOW) using computer classifier, as well as the graph of oriented gradient (hog (HOG)/pyramid oriented gradients (hog (Possible link) with computer classifier were created and evaluated in this study. Techniques from the categorization learning programmes were used to categorize hairy scalp images. Machine learning can achieve a reliability of 89.77 % when the training error is 1 104, which is much greater than that of the accuracy obtained by BOW + Vms (80.50 percent) and PHOG with SVM (80.50 percent) (53.0 percent) [7]

The goal of this study was to look at the current and potential therapeutic application of ai in hair regeneration and scalp problem diagnostics. Learned in the classroom solutions for head diagnoses and automatic hair loss evaluations, as well as soul devices, have indeed been suggested and explored. Hair restoration specialists ought to be aware of the advantages and limitations of these technological innovations when they become increasingly commonly accessible to physicians and patients.[8]

Researchers developed algorithms that successfully recognises and tallies particles and estimates hair length. Since our technique is based on a modular deep neural networks design, it may be modified to assess more endpoints, such as hair width. The usage of machine learning techniques has helped to improve patient care, study and enterprises offering hair development and wig retention goods[9] .

Researchers distinguished between dandruff and psoriasis using mobile phone hyperspectral imaging and analytics. The responsiveness, selectivity, and attentiveness of the machine learning (Replied, SVM, and MLP) for spectral classification were 65, 75, and 75, respectively. The sensitivities of the conventional techniques (ED and SAM) were 70 and 75, correspondingly. [10].

Figure 2 : Alopecia Areata

2.4 TECHNOLOGICAL DECISIONS

<

2.4.1 SUMMARY

The section outlines all resources and software that have been deployed to execute the project. Given the fact that there are several solutions on the market, some products outperform all competitors for the situation at hand.

2.4.2 PYTHON

Python is indeed an interpretive dialect that is very efficient in solving problems generally. This is one of the most widely utilized compute - intensive tools. It could really handle a wide range of tasks, such data gathering & extraction, and also program development. Python is indeed the main language of programming used in this work.

2.4.3 ANACONDA

Anaconda is a well-known framework that combines customers' access to nearly all the software and tools required to complete a Computer Analytics or Deep Learning assignment. This incorporates a range of programming environments. Also it comes with a variety of Libraries to work with.

2.4.4 SCIKIT-IMAGE

Scikit-image is a collection containing vision based & machine vision algorithms.

2.4.5 MATH

This mathematics package is really a basic Python package which is always available. It gives access to the functions of a core C programming language.

2.5 DATASET RESEARCH

2.5.1 DESCRIPTION

Machine learning models demand a large quantity of data or dataset to also be constructed upon successfully. This database should be thoroughly researched prior to beginning the project task. Our first task was to identify relevant datasets from which to build future trends. All main data for this project have indeed been taken from the disease dataset(Alopecia Areata, Melanoma, Tinea Capitis and Healthy scalp images). All datasets contain images, for disease detection. Following gathering data, then work for building models on it can start.

2.5.2 TINEA CAPITIS DATASET

Tinea Capitis images are collected from multiple sources. A total of 47 images were collected.

Figure 3 :Tinea capitis

2.5.3 ALOPECIA AREATA DATASET

Alopecia Areata images are collected from multiple sources. A total of 46 images were collected.

Figure 4 :Alopecia Areata

2.5.4 MELANOMA DATASET

Melanoma images are collected from multiple sources. A total of 48 images were collected.

Figure 5 :Melanoma

2.5.5 HEALTHY SCALP DATASET

Healthy scalp images are collected from multiple sources. A total of 22 images were collected.

Figure 6 : Healthy scalp

One of the most challenging parts of working with a collection has been its vast size. To eliminate the problem with extensive coding plus long delay, a majority of work had to be executed on its own aggregation of sub - sets.

2.6 CONCLUSION

Using well-known Image Processing methods and Machine Learning algorithms, this research seeks to detect disorders such as Tinea Capitis, Alopecia Areata, and Melanom The photos are processed in a pipeline, where they are first pre-processed and segmented, and then various characteristics are extracted in the next module.

Chapter 3

MACHINE LEARNING ALGORITHMS

3.1 OVERVIEW

Ideas and breakthrough research have only been accomplished over the last few decades. That amount of information generated for mankind has steadily increased. It really has been difficult for people to handle & analyze massive amounts of data because they have evolved from a limited data phase to an excessive one.

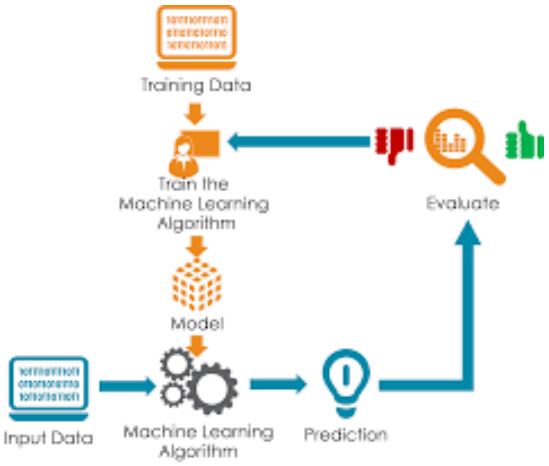


Figure 7 :ML Model

Machine learning has evolved throughout the order of popularity amid students. This has evolved beyond conceptual academia into economic research and applications. Through constantly emerging algorithms and systems that analyze huge amounts of data including complex calculations, benefits were gained. Machine learning is raising its standard using newly designed and optimized hardware available now.

3.2 SUPPORT VECTOR MACHINE

3.2.1 OVERVIEW

SVM is a supervised learning computing technique that can label things based on experience. Support vector machines are also widely employed in a growing number of scientific applications. SVM is, in theory, a mathematical entity, a strategy for maximizing a specified functional form in relation to a given data.

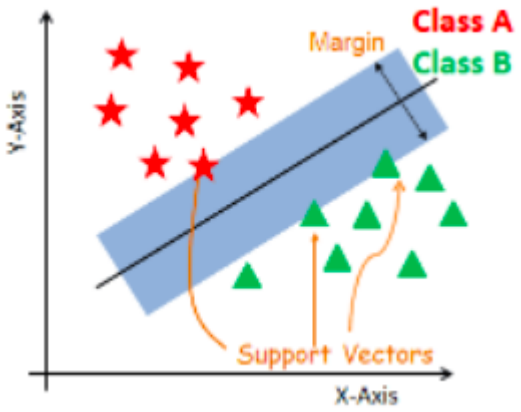


Figure 8 : SVM

3.2.2 ARCHITECTURE

SVMs, for the first estimate, create a separator line among data from two categories. Support vector machine is a method that generates data as an input then outputs a path which, when feasible, divides those classes. We make use of SVM technique for identifying the points both from classes that really are closest to the line. These are closely interrelated vectors. A proximity between both the lines and the support vectors is now computed.Our goal is to increase the profitability as much as achievable. The optimal hyper - plane is the one with the largest range.

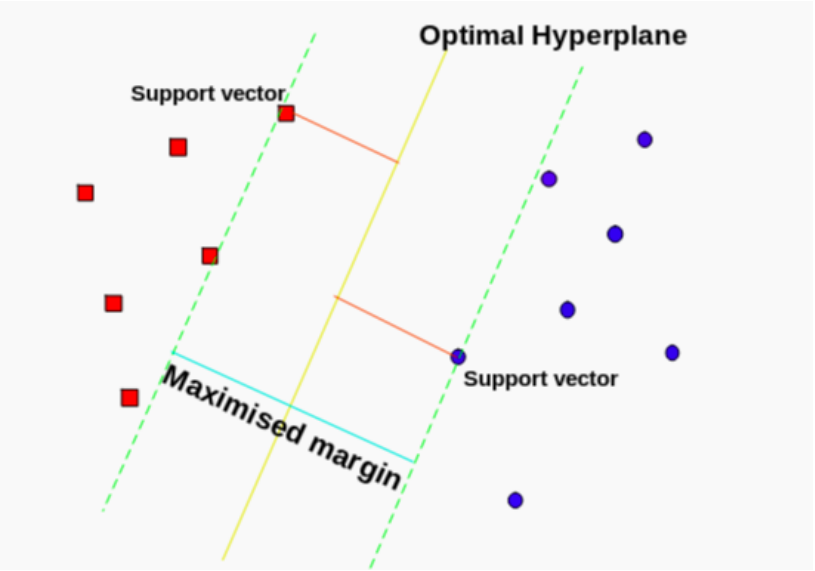


Figure 9: Optimal Hyperplane using the SVM algorithm

3.2.3 MATHEMATICAL EQUATION

Let us define three coordinates namely *a,b and c* where c has been defined with a constraint

$$c = a^2 + b^2$$

Thus, *c* is the square of the point's distance from the origin.

Let *d=c* that separates the data in high dimensional, wherein d is fixed.

$$\text{Since } c = a^2 + b^2, \text{ we derive } a^2 + b^2 = d$$

Which would be a circular expression.

Using this transformation, researchers can reflect the horizontal divider into extra dimensionality returned to its original dimensions.

3.3 LOGISTIC REGRESSION

3.3.1 OVERVIEW

The form of statistical analysis is frequently used for predictive modeling and forecasting, and also has ML algorithms. It has been used in statistical packages to calculate possibilities to analyze the overall relationship between the variables but one or even more relationships between the independent variables that used LR.

Within healthcare, this data analysis approach may be used to anticipate the likelihood for infection in a particular population, enabling again for deployment of precautionary care.

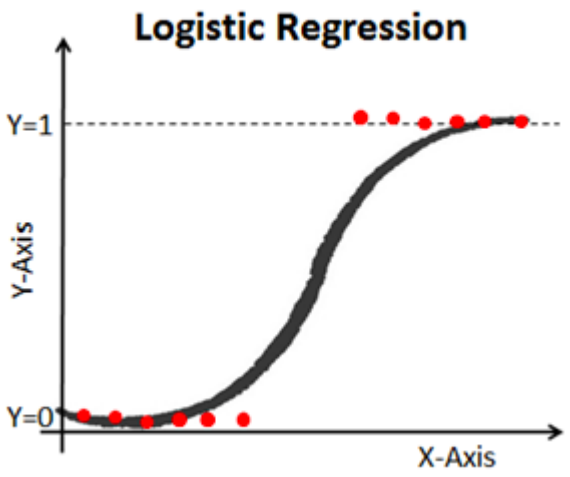


Figure 10 : LR

3.3.2 ARCHITECTURE

LR is a smart classification algorithm now at essence. For such a given set of features (or input), Y, a specific value (as well as outputs), x, could only accept distinct values inside a classification problem.

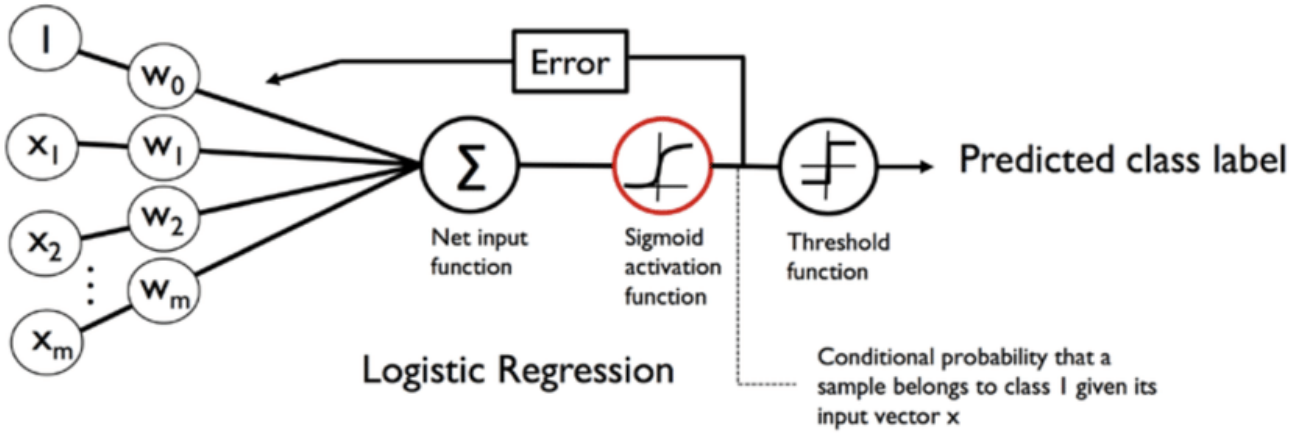


Figure 11 : LR Architecture

3.3.3 MATHEMATICAL EQUATION

The curve uses the variable A(independent) and relates to the mean that is rolling, B().

$$B = e^{c+dA} / 1 + e^{c+dA}$$

wherein B is the total probability of a 1, e is the base of the natural logarithm where c and d are the model parameters.

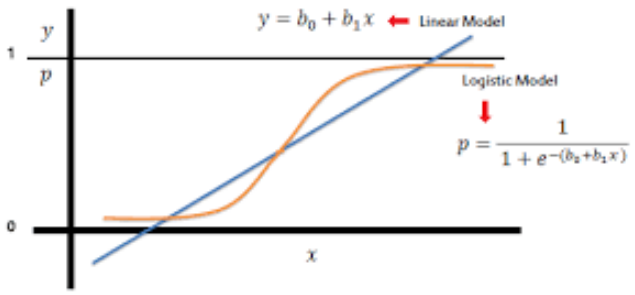


Figure 12 : LR math model

3.4 KNN

3.4.1 OVERVIEW

The k-nearest neighbors algorithm is indeed a non-parametric guided learning method. The output of a k-NN categorization is identified. The majority voting of its own neighbors defines an object, with the item assigned to a category more popular amongst k nearest neighbors . If k Equals one, then the entity was assigned to the class of an entity's closest neighbor.

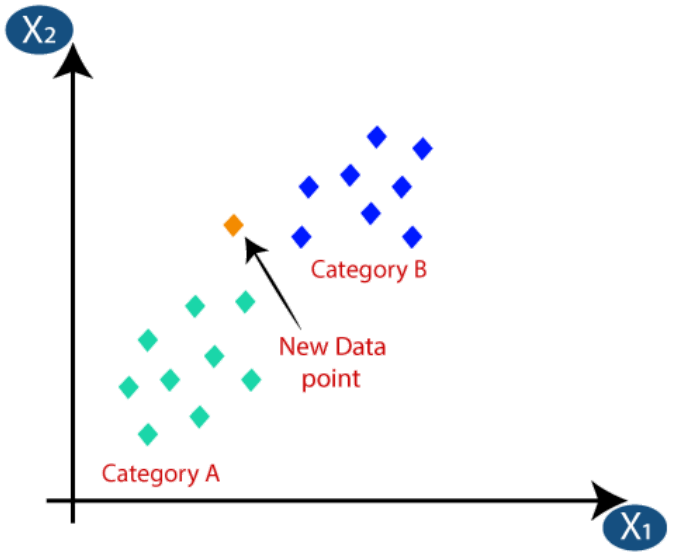


Figure 13 : KNN model

3.4.2 ARCHITECTURE

The algorithm's training phase consists simply in preserving both vectors and corresponding labels of training samples. The variable k is indeed a consumer variable mostly in the classification stage, as well as an unnamed vector is classified by assigning the name which is most frequent among k training images nearest to that query instance.

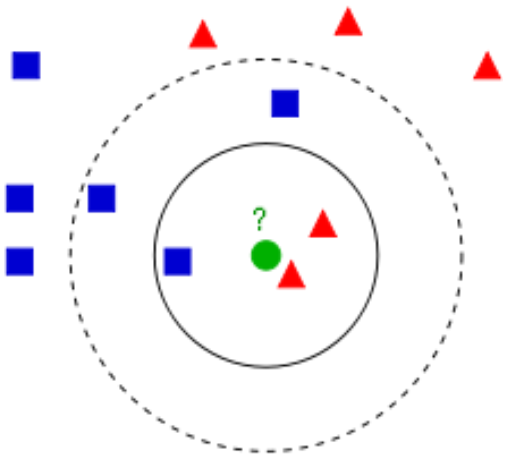


Figure 14 : KNN Architecture

3.4.3 ARCHITECTURE

k-nn classifier assigns a weight of 1/k to the k neighbors as well as a value of Zero to others. This is relevant to the scaled nearest neighbor classifier.

The optimal weighting scheme $\{w_{ni}^*\}_{i=1}^n$, that balances the two terms in the display above, is given as follows: set $k^* = \lfloor Bn^{\frac{4}{d+4}} \rfloor$,

$$w_{ni}^* = \frac{1}{k^*}\left[1 + \frac{d}{2} - \frac{d}{2k^{*2/d}}\{i^{1+2/d} - (i-1)^{1+2/d}\}\right] \text{ for } i = 1, 2, \ldots, k^* \text{ and}$$

$$w_{ni}^* = 0 \text{ for } i = k^* + 1, \ldots, n.$$

3.5 FINAL REMARKS

With any of these perspectives in mind, one may have a better understanding of what is happening in a Machine Learning model throughout its learning phase. There is also some uncertainty when developing ML models, however these ideas provide a better grasp of what might be happening if a system does not perform as predicted. In summary, understanding these ideas helps with code rectification.

Chapter 4

METHODOLOGY AND APPROACH

4.1 OVERVIEW

The two proposed methodologies of continuing forward with this project include Agile Process & CRISP-DM. Such techniques are adopted to aid in the development, the programming and the accomplishment for information retrieval objectives.

Agile technique allows for rapid iteration that adjusts to the demands of the user. Agile has aided with appropriately setting priorities throughout the order of priority in this project to guarantee that the project's primary goals are reached.

CRISP-DM is indeed a concept that is based just on machine learning model's development and advancement, because it provides a better knowledge about how to grow with this model using the moderate objective ladder listed below.

4.2 AGILE METHODOLOGY

According to the Agile Principles, Agile embraces four basic principles:

1. People & Connections emerge foremost, following procedures & technological methods.
2. functional technology upon top with comprehensive certifications & paperwork.
3. Strong cooperation between users and customers through contractual management.
4. Reacting with change instead of sticking to a fixed strategy.

Rather than formal licenses, the Agile Method focuses concentration upon customers. With agile, activities & goals were divided in 'customer tales,' where the workforce plans ahead again for work ahead & estimates how long it takes to complete the entire project.

A narrative could involve no coding by function, including study, discovery, and design, or it might be coding by essence, also including code rewriting or knowledge in new and component development. Those storylines then are divided in a sprint, every spanning 2 weeks. The team will work closely together and agree on objectives to also be completed during each iteration. In order to come up with a more thorough plan regarding what everyone is doing, the group will hold a team meeting to discuss the work done in the past week.

The questions can be the following:

1. What did they do that week ?
2. Their plan of action for that day?
3. If there are any obstacles they are facing?
4. What were the successful parts of the day?
5. What went wrong?
6. If they were able to rectify it?
7. What could be done to improve the existing system?

This system aids in the provision of maintenance.

4.3 CRISP-DM

CRISP-DM is a well-documented & openly accessible data gathering methodology. This provides the organization with elevated resources, applications, including technology which inspires better performance and provides organizations with architecture & configurations required to achieve faster and better data gathering results and outcomes. The below figure illustrates all six phases of a CRISP-DM model. CRISP-DM is indeed the ideal approach for data analytics, data gathering, & information training applications.

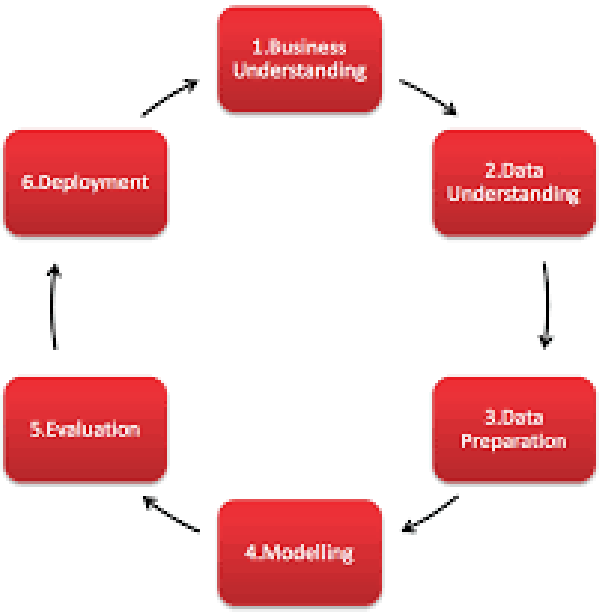


Figure 15 : CRISP DM

4.3.1 BUSINESS UNDERSTANDING

It is the most vital step because it deals with data gathering issue goals for complementing the business perception. The process includes understanding from a corporate and business standpoint, evaluating its company performance measure, then reviewing its company position in terms of holdings, money, skills, needs, dangers, expenses, and advantages. For the sake of this research, it involves determining how such a LDWS will create revenues for just a company. It necessitates the development of a project strategy in which the information learning goals are linked with business goals.

4.3.2 DATA UNDERSTANDING

Gather information, getting familiar with that as well, then eventually evaluating potential advantages, opportunities, & flaws all are part of this stage. This really is due to the fact that information doesn't really always match to a problem being solved. An analyzer and information prediction should comprehend that information's architecture, being able to define it, then finally confirm that information's integrity whilst maintaining uniformity in view. "Is the information full, do some lacking sections occur?" "Are there any blank result cells?" or "Has the data been investigated thoroughly?" are some of the frequent questions asked to a researcher.

4.3.3 DATA PREPARATION

All tasks associated with the obtained dataset are included in the Data Preprocessing step. It is a vital step prior to actually representing data, as well as the key steps in this phase have been: choosing this same information that would be most essential to a data gathering aim, trying to clean this same data and ensure quality of the data as well as rightness, changing things as well as constructing this same information besides changing the data in a fresh manner that aligns this same prototype to use for training, as well as incorporating this same information.

4.3.4 MODELING

This modeling stage entails choosing a model to aid in problem solving, including a regression analysis or even a basic classification algorithm, developing the model's architecture, using or adjusting various optimum values, then assessing the prototype model via calculating correctness and margin of error.

4.3.5 EVALUATION

This concept should be studied and evaluated before it is applied. The stage entails evaluating the models to check whether it can respond to and function with a strategic plan. At this point, it is vital to determine whether the model flourishes & thrives inside the software industry, nor whether the enterprise can meet the timelines and budget. It also should be determined whether the current proposal and model can meet the goals of the business challenge.

4.3.6 DEPLOYMENT

This distribution step consisted of several processes and is not often a data mining endeavor. It might denote a range of distribution activities, including a real application of this method to a client, planning its inspection or distribution, creating a final sentence, and reevaluating or reviewing a program. It could also be modified by the company problem's goals and objectives.

4.4 CONCLUSION

These 2 strategies employed enable a program's goals to also be divided and classified in numerous key stages. Agile divides initiatives in achievable , adaptable tasks, while CRISP-DM provides for just a greater understanding and comprehension of data analysis. The following section would go over the program's architecture & modeling components.

Chapter 5

SYSTEM DESIGN

5.1 OVERVIEW

This chapter identifies key design concepts for such work. Machine learning & information retrieval are important components of this quest. This developer's layout is critical to maintaining potential clients. This section will look at documents that pertain to something like a software's schematic implementation. Those architectural products were created as a result of different interactions with medical professionals and researchers.

5.2 DEFINITION, ACRONYMS AND ABBREVIATIONS

- ML- Machine Learning
- SVM- Support Vector Machine
- LR- Logistic Regression
- KNN

5.2.3 USE CASE DIAGRAM

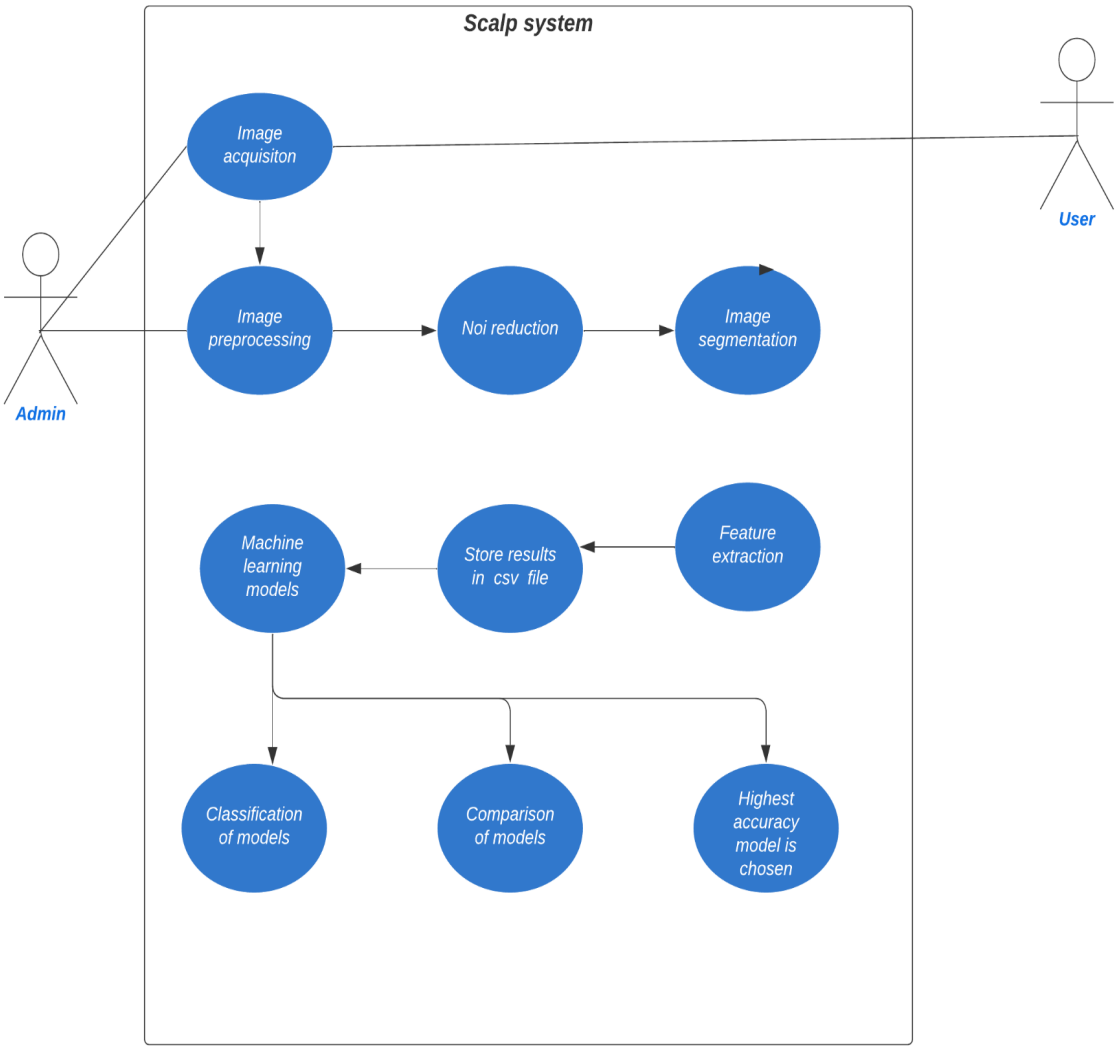


Figure 16 : Use case diagram for overall model

5.2.3.1 USE CASE DIAGRAM

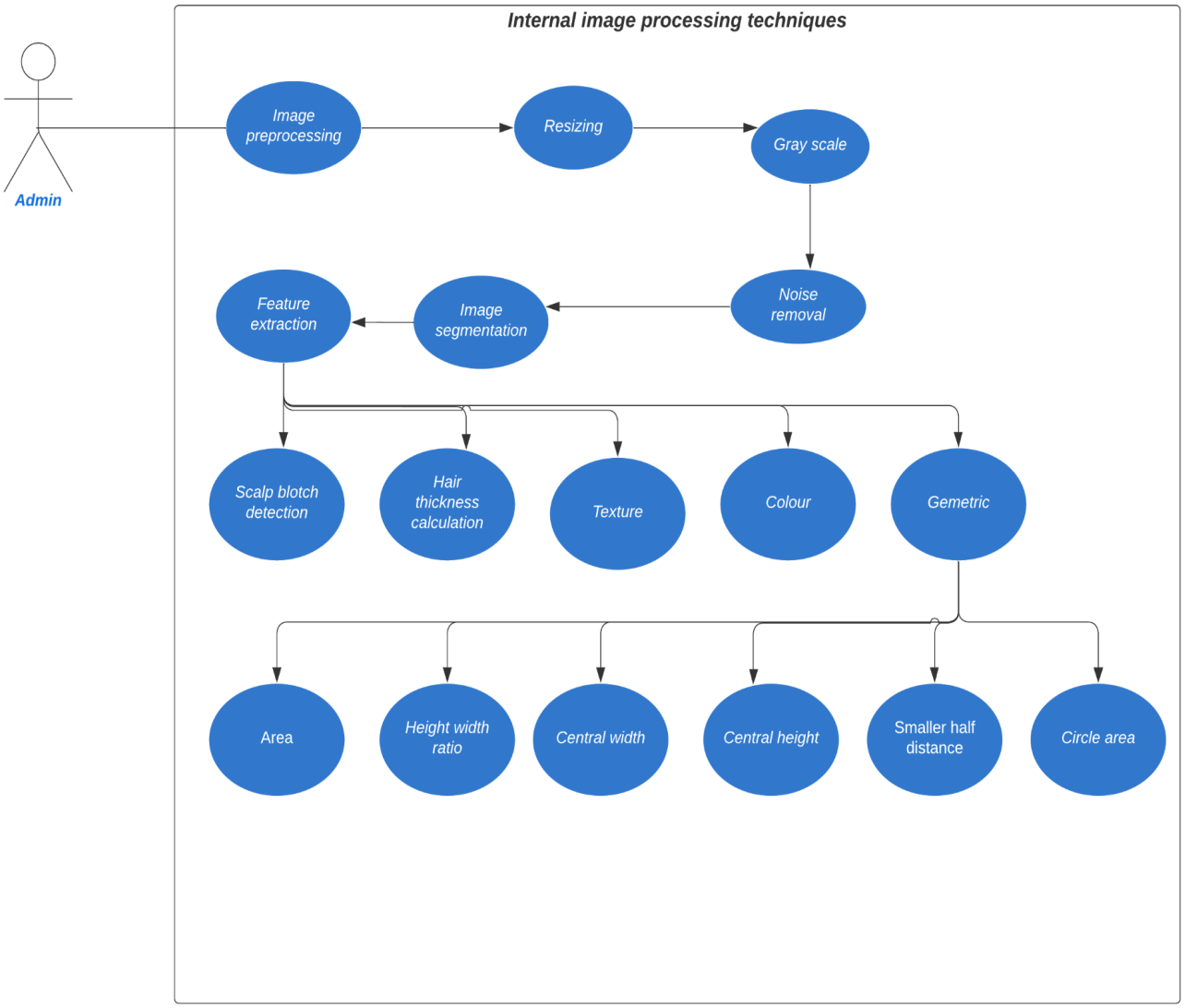


Figure 17 : Use case diagram for Image processing

5.2.4 CLASS DIAGRAM

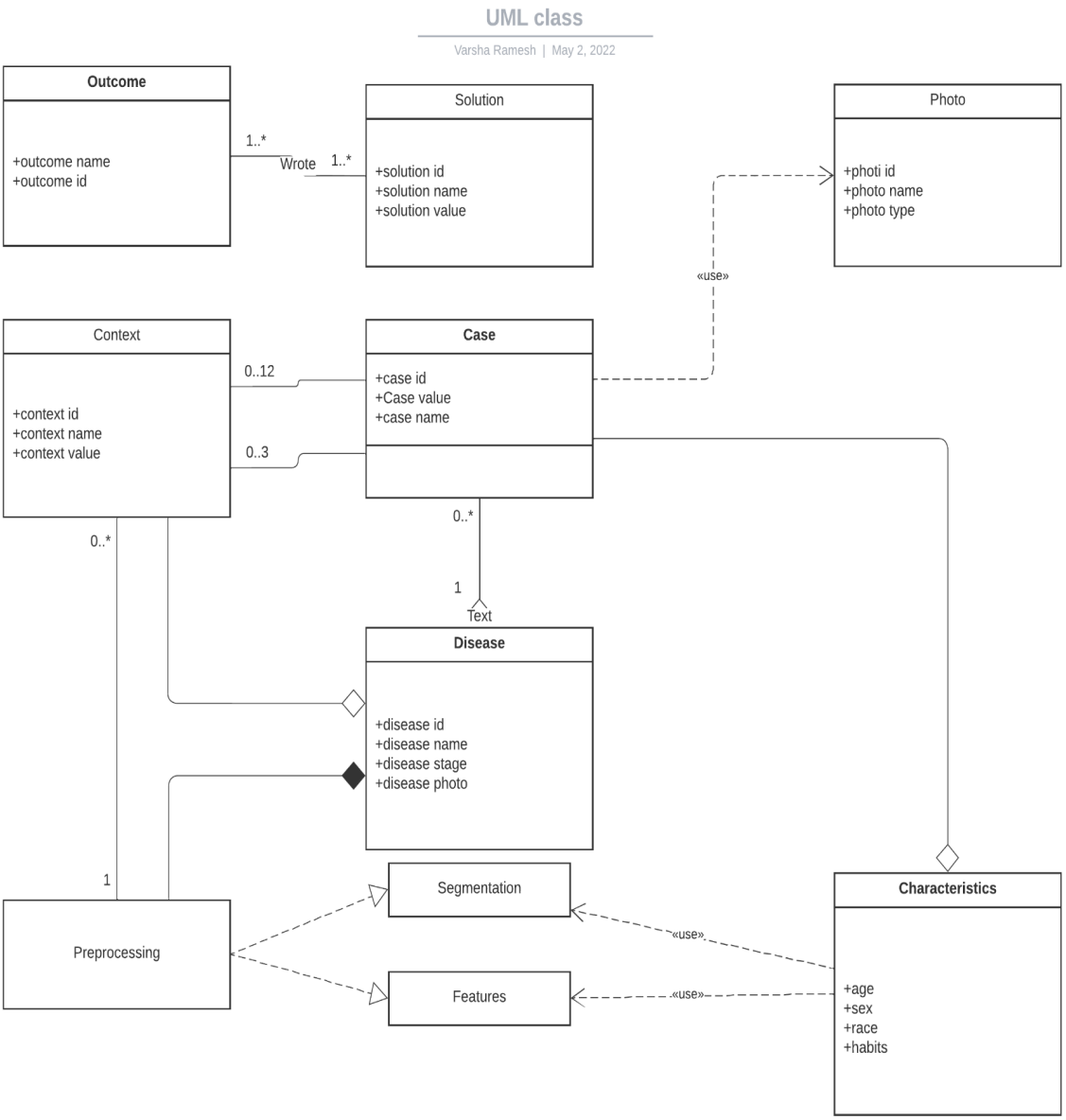


Figure 18 : Class diagram

5.2.5 SEQUENCE DIAGRAM

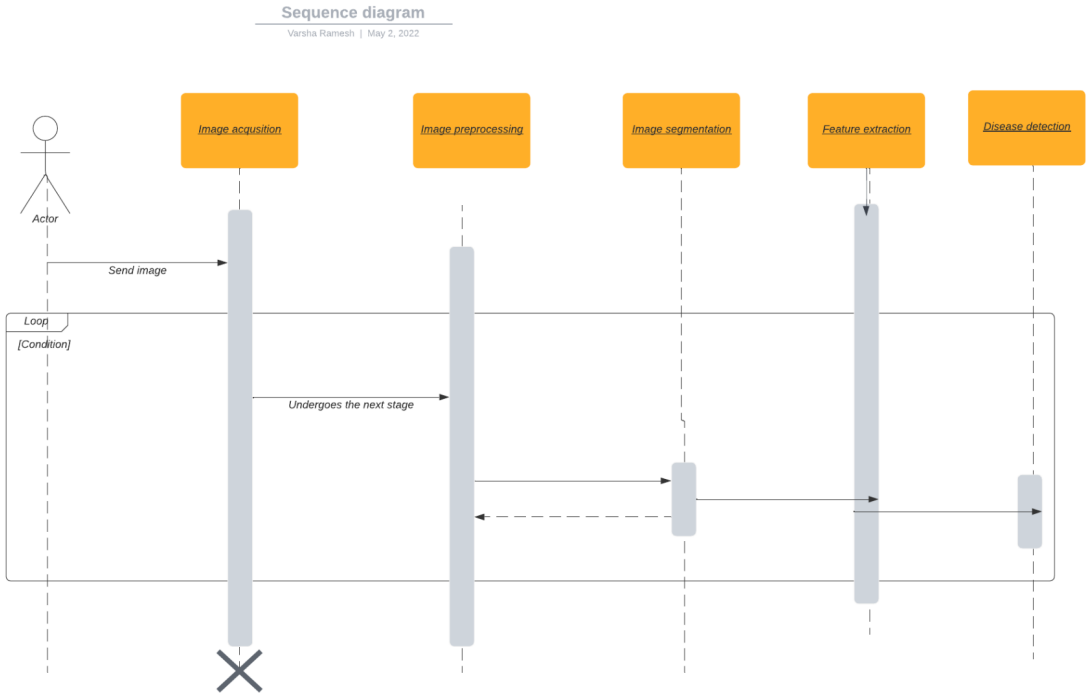


Figure 19 : Sequence diagram

5.2.6 ACTIVITY DIAGRAM

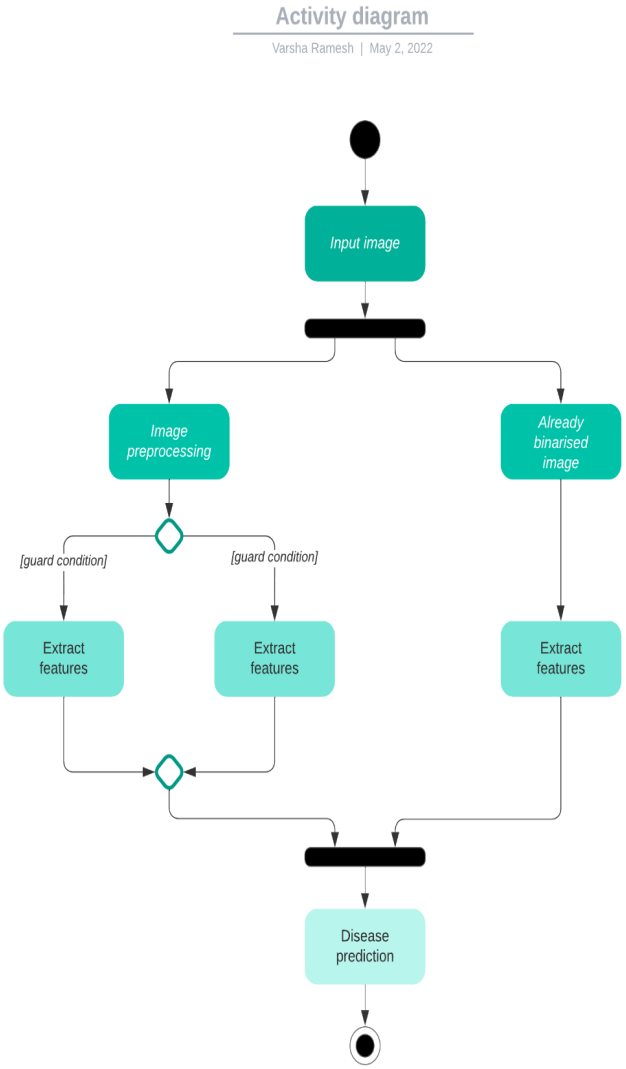


Figure 20 : Activity diagram

5.3 FINAL REMARKS

That chapter offers a summary of the concept documents included in this program; using these objects, it'll be simpler to define a program's human or engineering properties prior proceeding towards the implementation level. This validates that now the development adheres to just an agile model toward the fullest extent feasible, focusing just on customers than about the technical details of a software system.

The following section goes through the program's systems development.

Chapter 6

IMPLEMENTATION

6.1 OVERVIEW

- The phase of the process discusses what will be put into effect. This part brings the both data gathering & program innovation process to a close.

- Figure x depicts a task diagram that separates key goals of the scalp detection project.

- Figure y depicts a production process flow that separates main goals of a disease categorization project.

- Specific Loading and Data Interpretation courses were influenced by numerous publications through disease prediction

- The primary purpose among these mining techniques is to construct a machine learning method that includes an image as input and determines whether it's a diseased scalp or not.

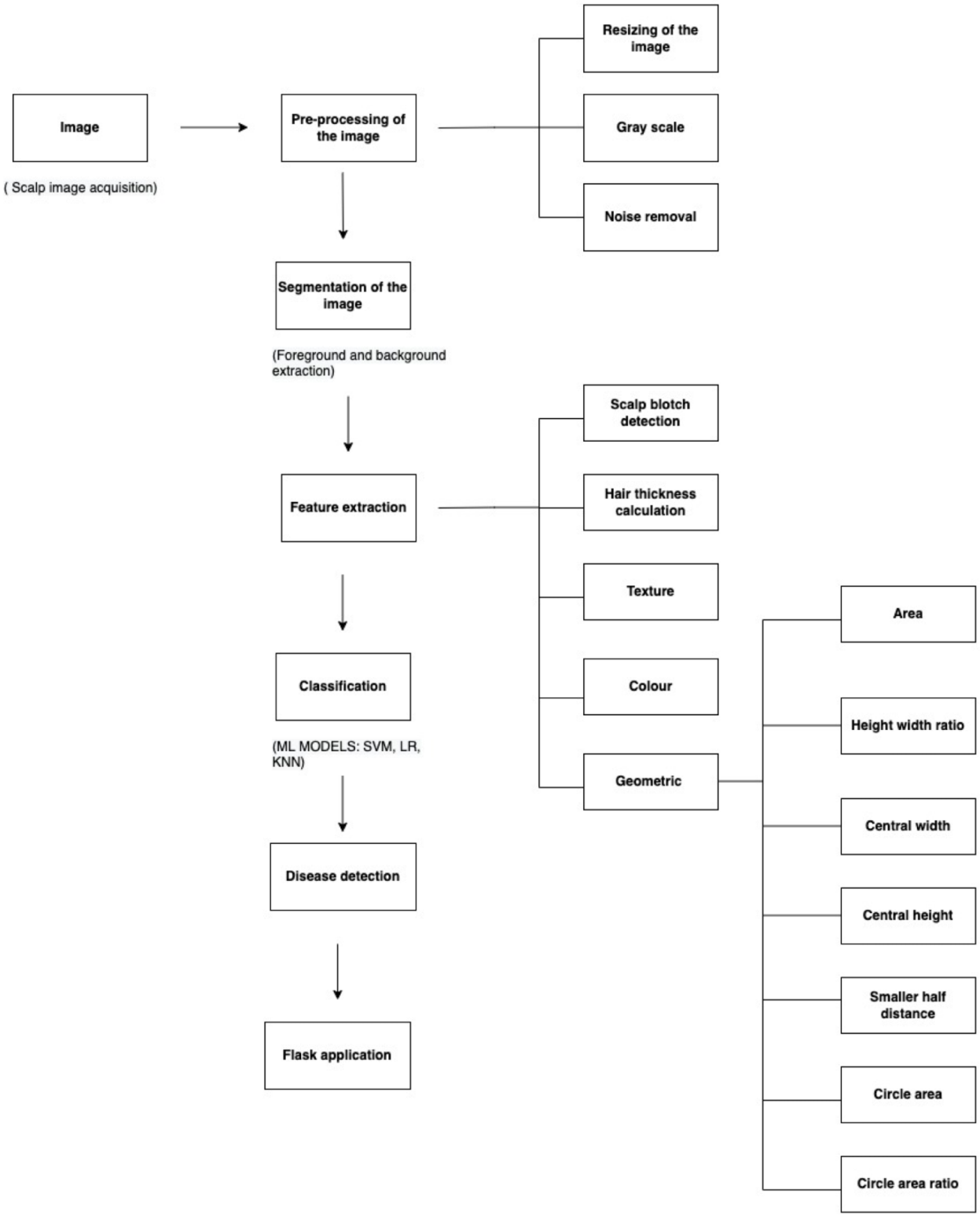


Figure 21 : Flowchart

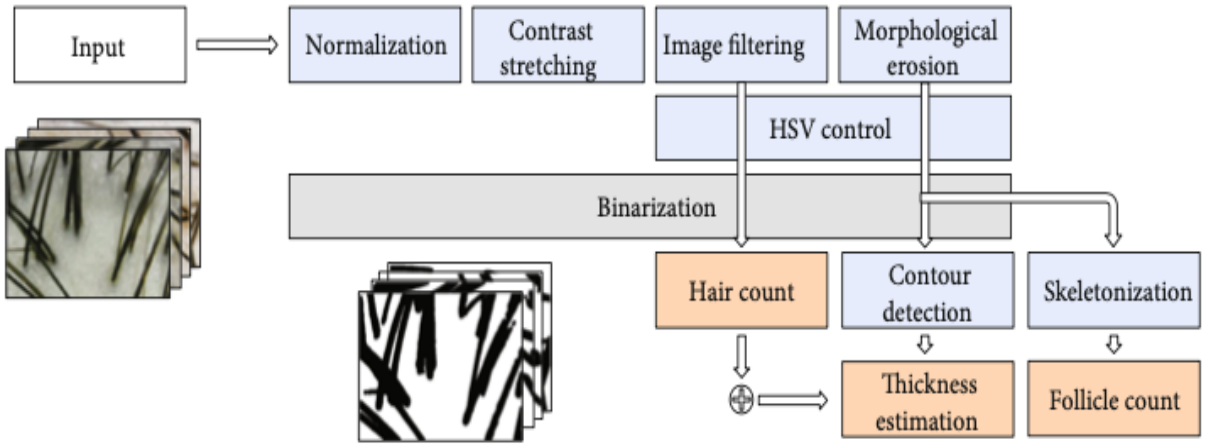


Figure 22 : Processes of image processing

6.2 SCALP DISEASE DETECTION

6.2.1 GENERAL PROCEDURE

The figure below depicts the overall general concept for the freshly designed Scalp model. To begin, the whole first scalp image must be normalized via monochrome preparation, pattern classification, then image softening in that order. Histogram management is a prerequisite for thresholding, and binary image implies removing ROI from the basis. Photo smoothness seeks to eliminate the picture's worthlessness or failure to identify, such as turbulence. The ability of ROI is then used for the route search for meaning to extract the portion of the image that includes path lines. From there, they use pattern identification to extract all boundaries of the pathways in ROI.

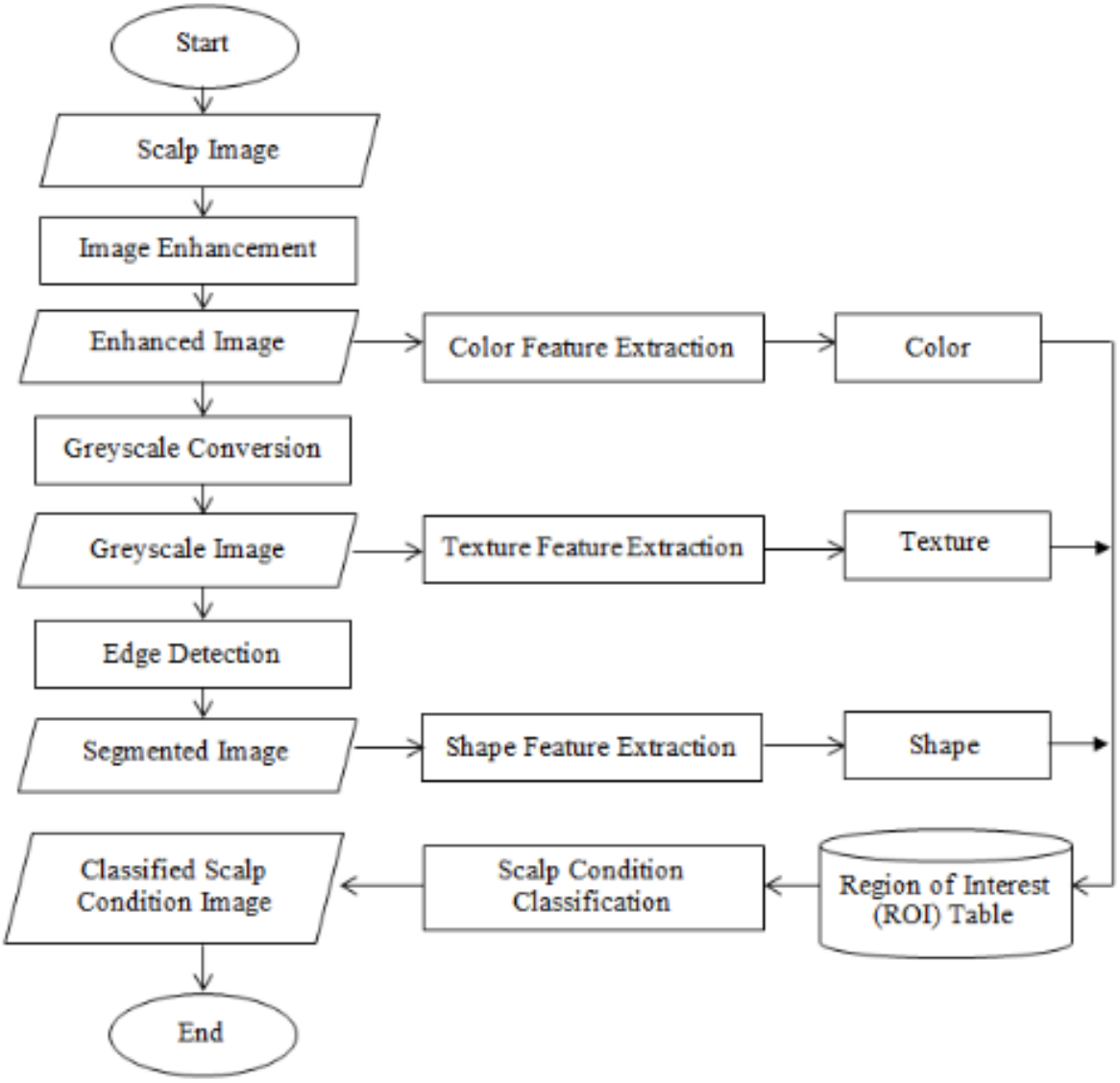


Figure 23 :Feature extraction

6.2.2 IMAGE PROCESSING

The obtained photos are subsequently passed through a series of processes using machine learning such as feature extraction, which describe the characteristics of significant elements of each sign. Since users concentrate solely on the scalp, the artificial intelligence also aids inside the elimination from backdrop items which may interfere with the right recognition of the sign. It aids in the elimination of noise to a larger level.

Figure 24 :Dataset

6.2.2.1 GRAYSCALE PROCESSING

One objective of monochrome preparation would be to convert the initial scalp image, that is in RGB coloring, to a gray level. In the RGB coloring area, the orientations for each pixel are 'a','b','c' where they represent red, green and blue. The monochrome management is the ongoing process of converting 'a','b' and 'c' to a dimmer number in the range from 0 to 255. A color of these pixels would generally remain white even as dim quality.

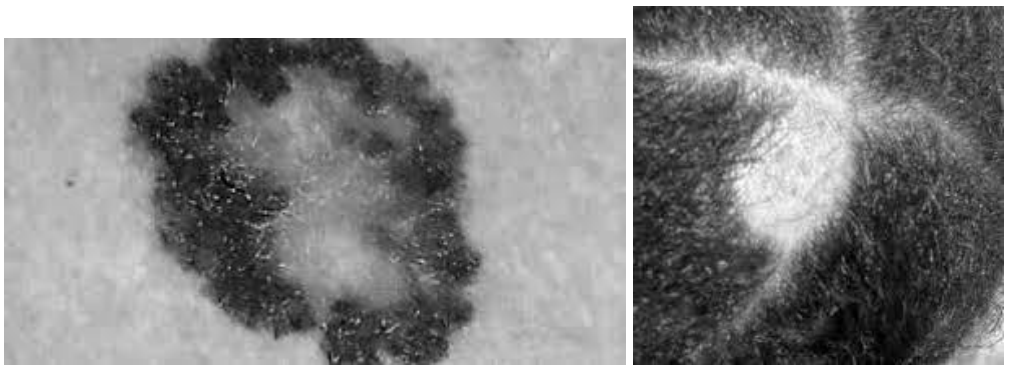


Figure 25 :Grayscale xxxxiv

6.2.2.2 IMAGE SEGMENTATION

We adopt the GrabCut method, dividing the picture's current frame separation. A Gmm method is being used to categorize & aggregate particles inside the selected area by analyzing both depth of field.

For the main reason: Because of unevenness of the backdrop, the GrabCut algorithm is preferred over the other approaches. This method operates quite well in a situation because of the error model, which states that the locations or wavelengths of a message will not be established simultaneously.

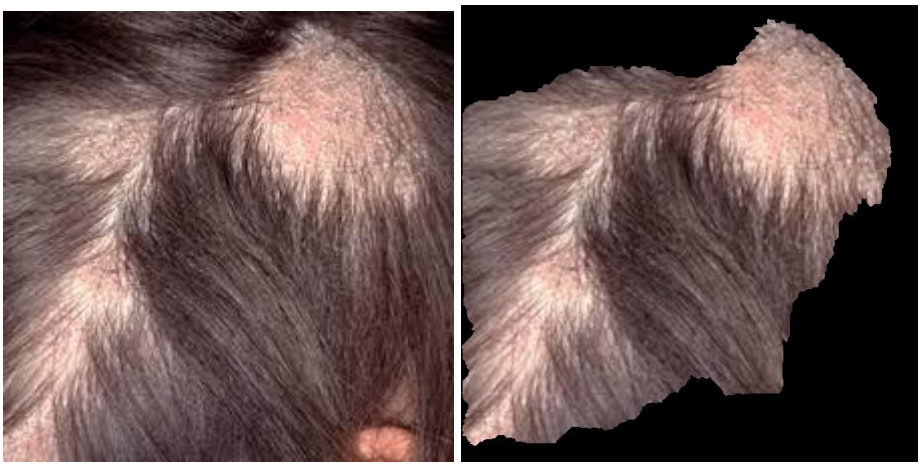


Figure 26 :Segmented image

6.2.3 FEATURE EXTRACTION

These photos were analyzed inside a workflow, where they can be firstly cleaned and fragmented, and then various aspects are retrieved. The retrieved characteristics are utilized to create classification techniques and diagnose illnesses within the scope of this report.

6.2.3.1 CHROMATIC FEATURES

Red-green-blue values for four regularly seen scalp colors are established as measured value whereby the frames in the picture will be compared. The Manhattan distance is computed between both the RGB Red-green-blue values of the pixel as well as the baseline color. As a result, the shade of a picture is determined as the standard tone wherein the bulk of an original image has a minimal distance.

The defined pixel values were:

Black - [150 , 80 , 121]

White - [65 , 60 , 91]

Peach - [50 , 65 , 146]

Yellow - [80 , 90 , 100]

Figure 27 :Color feature

6.2.3.2 GEOMETRIC FEATURES

By conjunction to that same color feature, the pictures are retrieved with other template matching such as lengths & related proportions.

1. Area : This amount of non-monochromatic pixels is used to compute its surface.

2. Altitude proportion : That's the proportion of the scalps height (h) towards its width (w).

3. Width from the middle : The distance from either the current node in the furthest place just on x axes is used to determine breadth.

4. Height from the middle : A dividing line from the pixel intensity to the far location just on y axes is used to calculate the elevation.

5. Shorter ½ length : The shorter halves length is indeed the halfway spacing of both the altitude or even the width, and this is determined based upon what is lower.

6. Round diameter : It's also described as the radius visible on the scalp using the shorter 1/2 distance calculation.

7. Round diameter ratio : A circular size factor is the proportion of a concentric circle to the length of the scalp.

6.2.3.3 SKELETONISATION

It is the technique of decreasing front portions inside a monochrome picture to a skeleton remnant which retains a previous area's size & connection whilst discarding the majority of an underlying pixel intensities.

It keeps its topological (maintains its initial item's structure), it maintains the form (a significant characteristic appropriate in object identification or categorization is retrieved), it causes the "skeletal" to remain in the center of the image.

Then obtain that picture's geometry with numpy.zeros and retrieve its transform by applying a filter to the picture's form; with that picture, we retrieve a segmentation technique that uses a morphing cross. Then stretch, degrade, and reduce the image. The skeleton is then obtained by doing binary arithmetic on the picture.

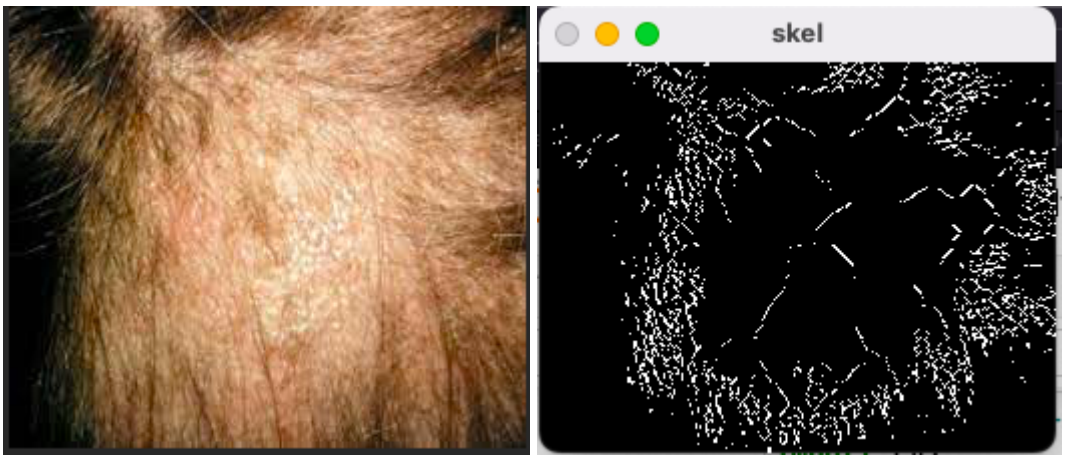


Figure 28 :Skeletonisation

```
50310
```

6.2.3.4 HAIR COUNT

Lines identification is indeed an image analysis method that utilizes a sample of 'm' terminal nodes then determines together all routes upon which the boundary endpoints reside. The Hough transform does have a benefit over connected layer techniques within these bits solely on a single line; it should not have to be fully continuous. If attempting to pinpoint arcs with small gaps within between excessive noise, or even when items are obscured, it can be very beneficial. The Hough transform is indeed a methodology for extracting features from images which is used in image classification, machine learning, and image recognition. The technique's goal is to use a voting mechanism to locate imperfect examples of objects inside a specific class of forms.

So first it builds a two - dimensional array which is originally established to zero. We use rows and columns to represent the array. The length of the matrix is determined by the level of precision required. It just provides a set of results. First, the source picture ought to be a scaled version, thus utilizing a cutoff before performing the hough transform. The 2nd and 3rd variables are accuracies. The final input is the criteria, which would be the lowest number of votes it requires consideration as just a line. As a result, it reflects the shortest amount of columns which should be identified.

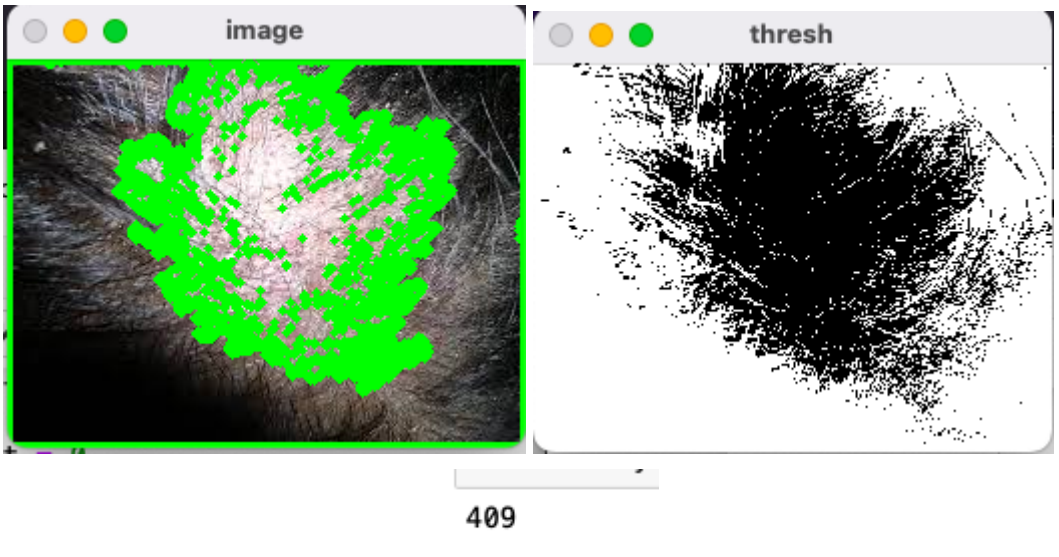


Figure 29 :Hair count

6.2.3.5 TEXTURAL FEATURES

The unprocessed photos are processed to remove relevant roughness, which will then be designed to recognize covering, smoothing, creases, and microscopic holes, which each represents a particular disease alteration in the body.

Count of curvatures: The amount of curves, as well as their region but per contoured region, were obtained by transforming the grayscale picture with such a Gaussian blur filter then using a thresholding technique to the dimension reduction. The outlines were then computed to use the bounding box photos, and a criterion was used to screen out the photographs that contained textured pictures.

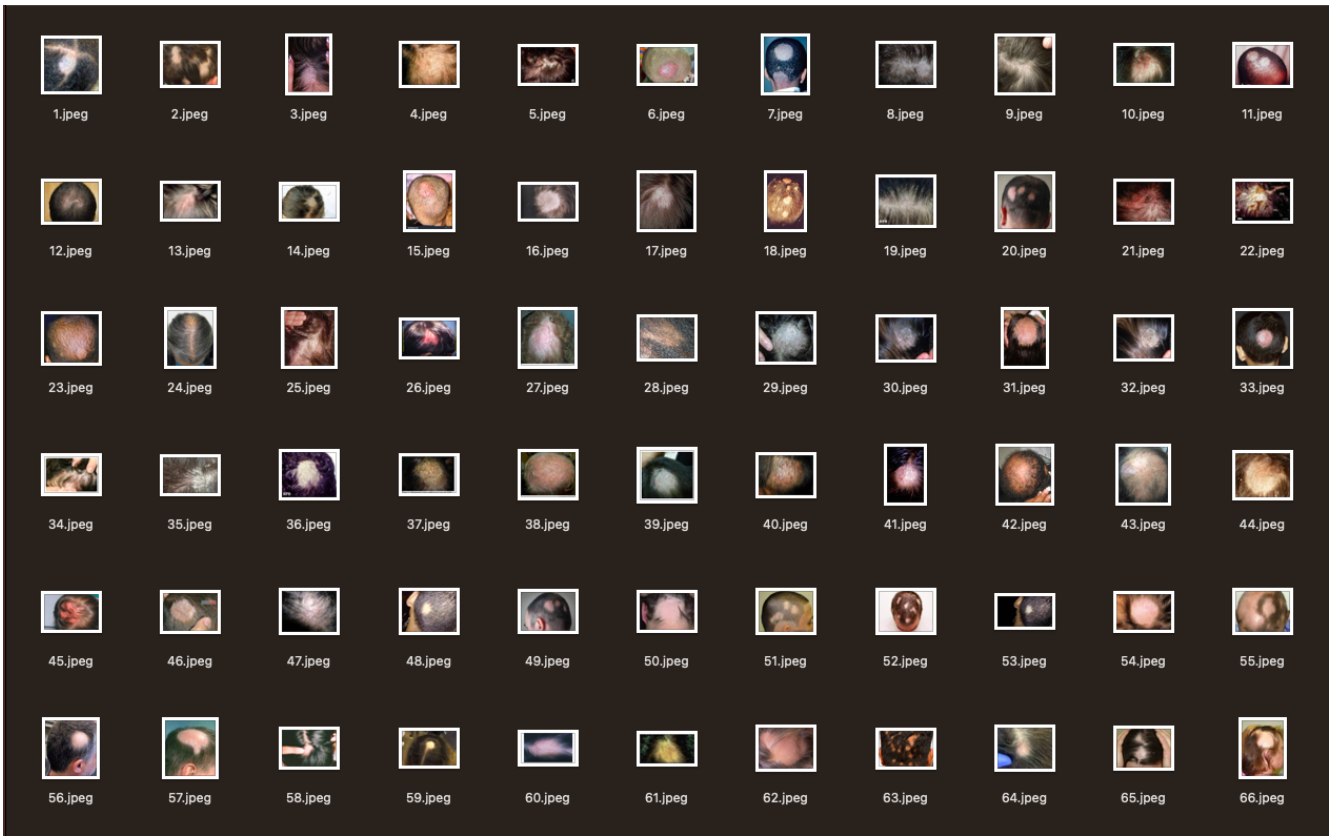
Length of a shape: When we have the amount of contour, then test to see if it is below than thousand and add it to the area.

Distance of a curve: When we have the surface of a curve, we utilize cv2.arclength() to determine overall distance of the curves then add it by increasing the sum to a duration of a curve.

Chapter 7

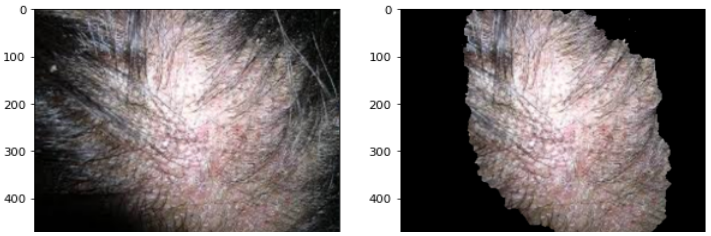
CODE AND RESULTS

7.1 DATASET



7.2 AFTER SEGMENTATION

In [74]:
```
#segmentation
img=data[46]
mask = np.zeros(img.shape[:2],np.uint8)

bgdModel = np.zeros((1,65),np.float64)
fgdModel = np.zeros((1,65),np.float64)

rect = (1,1,545,545) #rect = (start_x, start_y, width, height)

cv2.grabCut(img,mask,rect,bgdModel,fgdModel,5,cv2.GC_INIT_WITH_RECT)
mask2 = np.where((mask==2)|(mask==0),0,1).astype('uint8')
fimg = img*mask2[:,:,np.newaxis]


fig=plt.figure(figsize=(10,10))

ax1 = fig.add_subplot(121)
ax1.imshow(img)

ax2 = fig.add_subplot(122)
ax2.imshow(fimg)

#plt.colorbar()
plt.show()

fimg=cv2.cvtColor(fimg,cv2.COLOR_BGR2RGB)
cv2.imwrite('/Users/varsha/Desktop/Capstone project/segmented_1/s46.jpg', fimg)
```



7.3 COLOR FEATURE

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import cv2
from skimage.color import rgb2gray
from skimage.io import imread, imshow
from scipy import ndimage
from PIL import Image,ImageChops
#from colours import color_analysis1
import math
import os
import glob
import random


from collections import Counter
import webcolors
import cv2
from PIL import Image

df=pd.read_csv('/Users/varsha/Desktop/Capstone project/resp2.csv')
print(df.head())
list(df.columns)
df.shape

# COLOURS = {   "red": (255, 0, 0),
#               "green" : (0,255,0),
#               "blue":(0,0,255),
#               "pink":(255,192,203),
#               "white":(255,250,250),
#               "yellow":(255,255,0),
#               "purple":(221,160,221)
#               }

COLOURS = { "yellow": (51, 42, 31),
           "white":(65,50,90),
           "peach":(50,65,146),
           "black": (80,90,100)
          }
```

```
def classify(rgb_tuple):

    manhattan = lambda x,y : abs(x[0] - y[0]) + abs(x[1] - y[1]) + abs(x[2] - y[2])
    distances = {k: manhattan(v, rgb_tuple) for k, v in COLOURS.items()}
    color = min(distances, key=distances.get)
    return color


def color_analysis(image):
    color_counter = Counter({color: 0 for color in Counter(COLOURS)})
    for pixel_count, RGB in image.getcolors(image.width * image.height):
        if(RGB!=(0,0,0)):
            color_name = classify(RGB)
            color_counter[color_name] += pixel_count


    for color in color_counter:
        pixel_count = image.width * image.height
        color_counter[color] = color_counter[color] / pixel_count


    colour = max(color_counter, key=color_counter.get)

    return colour

img_dir = "/Users/varsha/Desktop/Capstone project/segmented_2"
data_path = os.path.join(img_dir,'*g')
files = glob.glob(data_path)
```

```
img_dir = "/Users/varsha/Desktop/Capstone project/segmented_2"
data_path = os.path.join(img_dir,'*g')
files = glob.glob(data_path)

colour=[]
for i in range(47,94):
    for f1 in files:
        tmp=f1.split("/")
        a=tmp[-1].split("s")
        b=a[1].split(".")
        n=b[0]
        if(i==int(n)):
            img = Image.open(f1).convert('RGB')
            col=color_analysis(img)
            colour.append(col)

df['colour']=colour
print(df.head())

with pd.option_context('display.max_rows', None, 'display.max_columns', None):  # more options can be specified also
    print(df)

df.colour.unique()

df.to_csv (r'/Users/varsha/Desktop/Capstone project/2.csv', index = False, header=True)
```

	colour
0	peach
1	yellow
2	peach
3	peach
4	red
5	pink
6	peach
7	peach
8	yellow
9	peach
10	peach
11	yellow
12	peach
13	pink
14	peach
15	pink
16	peach
17	yellow

DISEASE	MEAN RED	MEAN GREEN	MEAN BLUE
TINEA CAPITIS	66.54	56.37	51.75
ALOPECIA AREATA	66.10	54.96	50.61
MELANOMA	90.29	69.12	61.18
HEALTHY SCALP	63.67	52.96	49.22

Table 1 : Inference for color feature

7.4 GEOMETRIC FEATURE

```
#Area
area=[]
for img in df['segmented_img']:
    cnt=0
    img=cv2.cvtColor(img,cv2.COLOR_BGR2GRAY)
    img=cv2.filter2D(img, -1, kernel)
    for i in range(0,img.shape[0]):
        for j in range(0,img.shape[0]):
            if(img[i][j]!=0):
                cnt+=1

    area.append(cnt)

df['area']=area
print(df.head())
```

```
central_width=[]
central_height=[]
height_width_ratio=[]
mid_x=280
mid_y=280
for img in df['segmented_img']:
    img=cv2.cvtColor(img,cv2.COLOR_BGR2GRAY)
    #img=cv2.filter2D(img, -1, kernel)
    left=0
    right=0
    top=0
    bot=0
    for i in range(mid_x,0,-1):
        if(img[mid_x][i]!=0):
            left+=1

    for j in range(mid_x,560):
        if(img[mid_x][j]!=0):
            right+=1

    for k in range(mid_y,0,-1):
        if(img[k][mid_y]!=0):
            top+=1

    for l in range(mid_y,560):
        if(img[l][mid_y]!=0):
            bot+=1

    h=top+bot
    w=left+right
    h_w=h/w
    height_width_ratio.append(h_w)
    central_width.append(w)
    central_height.append(h)

df['height_width_ratio']=height_width_ratio
df['central_width']=central_width
df['central_height']=central_height
print(df.head())
```

```
#Smaller-Half-distance, Circle Area, Square Area
smaller_half_dist=[]
circle_area=[]
circle_area_ratio=[]
square_area=[]
square_area_ratio=[]
small=0
circ_a=0

for a in df.index:
    small=min(df['central_width'][a],df['central_height'][a])
    shd=small/2
    circ_a=math.pi*(shd**2)
    circ_a_rat=circ_a/df['area'][a]
    square_a=4*(shd**2)
    square_a_rat=square_a/df['area'][a]
    square_area.append(square_a)
    square_area_ratio.append(square_a_rat)
    smaller_half_dist.append(shd)
    circle_area.append(circ_a)
    circle_area_ratio.append(circ_a_rat)


df['smaller_half_dist']=smaller_half_dist
df['circle_area']=circle_area
df['circle_area_ratio']=circle_area_ratio
print(df.head())
```

```
                                          segmented_img    area  \
0  [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], ...   98741
1  [[[0, 0, 2], [0, 0, 2], [0, 0, 2], [0, 0, 2], ...  167617
2  [[[1, 0, 0], [1, 0, 0], [1, 0, 0], [1, 0, 0], ...  223022
3  [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], ...  198454
4  [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], ...  148689

   height_width_ratio  central_width  central_height
0            1.515284            229             347
1            1.235437            412             509
2            0.902128            470             424
3            1.153318            437             504
4            0.844920            374             316
   ID  Tinea captisis  Alopecia areata  Melanoma  Healthy scalp     Img  \
0   1               1                0         0              0  1.jpeg
1   2               1                0         0              0  2.jpeg
2   3               1                0         0              0  3.jpeg
3   4               1                0         0              0  4.jpeg
4   5               1                0         0              0  5.jpeg

                                          segmented_img    area  \
0  [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], ...   98741
1  [[[0, 0, 2], [0, 0, 2], [0, 0, 2], [0, 0, 2], ...  167617
2  [[[1, 0, 0], [1, 0, 0], [1, 0, 0], [1, 0, 0], ...  223022
3  [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], ...  198454
4  [[[0, 0, 0], [0, 0, 0], [0, 0, 0], [0, 0, 0], ...  148689

   height_width_ratio  central_width  central_height  smaller_half_dist  \
0            1.515284            229             347              114.5
1            1.235437            412             509              206.0
2            0.902128            470             424              212.0
3            1.153318            437             504              218.5
4            0.844920            374             316              158.0

     circle_area  circle_area_ratio
0   41187.065087           0.417122
1  133316.625848           0.795365
2  141195.740223           0.633102
3  149986.701866           0.755776
4   78426.719004           0.527455
```

DISEASE	AREA	HEIGHT WIDTH RATIO	CENTRAL WIDTH	CENTRAL HEIGHT	SMALLER HALF DISTANCE	CIRCLE AREA	CIRCLE AREA RATI
TINEA CAPITIS	167961.4	1.0899	410.934	432.869	190.2	117723.9	0.686
ALOPECIA AREATA	158345.4	1.1987	370.217	417.847	171.6	98762.14	0.602
MELANOMA	184629	1.2065	392.270	449.791	185.6	114174.23	0.594
HEALTHY SCALP	186891.1	1.1986	408.793	468.275	191.7	118030.96	0.634

Table 2 : Inference for geometric feature xxxxxvii

7.5 SKELETONISATION FEATURE

```
In [*]: #skeletonization- just image- see if you should count lines with this
        import cv2
        import numpy as np


        img = cv2.imread('/Users/varsha/Desktop/Capstone project/images_1/4.jpeg',0)
        size = np.size(img)
        skel = np.zeros(img.shape,np.uint8)

        ret,img = cv2.threshold(img,127,255,0)
        element = cv2.getStructuringElement(cv2.MORPH_CROSS,(3,3))
        done = False

        while( not done):
            eroded = cv2.erode(img,element)
            temp = cv2.dilate(eroded,element)
            temp = cv2.subtract(img,temp)
            skel = cv2.bitwise_or(skel,temp)
            img = eroded.copy()
            #thinned = cv2.ximgproc.thinning(cv2.cvtColor(img, cv2.COLOR_RGB2GRAY))

            zeros = size - cv2.countNonZero(img)
            if zeros==size:
                done = True


        print(zeros)
        print(skel)
        cv2.imshow("skel",skel)
        cv2.waitKey(0)
        cv2.destroyAllWindows()
```

```
50310
```

DISEASE	SKELETONISATION
TINEA CAPITIS	43699.3
ALOPECIA AREATA	37759.5
MELANOMA	29527.1
HEALTHY SCALP	29542.2

Table 3 : Inference for skeletonisation

7.6 HAIR COUNT FEATURE

```
: #hough tranform to count the number of lines-check if you should with segmented/ normal okay
import cv2

lines = []
image = cv2.imread('/Users/varsha/Desktop/Capstone project/images_1/47.jpeg')
gray = cv2.cvtColor(image,cv2.COLOR_BGR2GRAY)
thresh = cv2.threshold(gray, 120, 255, cv2.THRESH_BINARY_INV)[1]

cnts = cv2.findContours(thresh, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
cnts = cnts[0] if len(cnts) == 2 else cnts[1]

lines = 0
for c in cnts:
    cv2.drawContours(image, [c], -1, (36,255,12), 3)
    lines += 1


df['lines']=lines
print(lines)
cv2.imshow('thresh', thresh)
cv2.imshow('image', image)
cv2.waitKey()
```

```
409
```

DISEASE	HAIR COUNT
TINEA CAPITIS	88.5
ALOPECIA AREATA	49.95
MELANOMA	119.43

Table 4 : Inference for hair count

7.7 TEXTURAL FEATURE

```
#Contours (Texture/Patches)
num_cont=[]
area_cont=[]
per_cont=[]
for i in range(2,91):
    for f1 in files:
        tmp=f1.split("/")
        a=tmp[-1].split("s")
        b=a[1].split(".")
        n=b[0]
        if(i==int(n)):
            img = cv2.imread(f1,1)
            gray = cv2.cvtColor(img,cv2.COLOR_RGB2GRAY)
            blur = cv2.GaussianBlur(gray, (3,3),0)
            thresh = cv2.adaptiveThreshold(blur, 255, cv2.ADAPTIVE_THRESH_GAUSSIAN_C, cv2.THRESH_BINARY_INV, 205, 1)
            contours, _ = cv2.findContours(thresh, cv2.RETR_LIST, cv2.CHAIN_APPROX_SIMPLE)
            filtered = []
            tot_are=0
            tot_per=0
            for c in contours:
                if cv2.contourArea(c) < 1000:
                    continue
                filtered.append(c)

            for c in filtered:
                area = cv2.contourArea(c)
                p = cv2.arcLength(c,True)
                tot_are+=area
                tot_per+=p


            num_cont.append(len(filtered))
            area_cont.append(tot_are)
            per_cont.append(tot_per)

df['num_contours']=num_cont
df['area_contours']=area_cont
df['len_contours']=per_cont
print(df.head())
```

```
   ID  Tinea captisis  Alopecia areata  Melanoma  Healthy scalp     Img  \
0   1               1                0         0              0  1.jpeg
1   2               1                0         0              0  2.jpeg
2   3               1                0         0              0  3.jpeg
3   4               1                0         0              0  4.jpeg
4   5               1                0         0              0  5.jpeg

                                                 pic  num_contours  \
0  [[[60, 67, 75], [60, 66, 75], [58, 65, 73], [5...             3
1  [[[92, 81, 77], [84, 73, 69], [66, 55, 51], [5...             6
2  [[[6, 58, 98], [6, 58, 98], [17, 69, 108], [28...            10
3  [[[4, 0, 0], [5, 1, 0], [7, 3, 0], [9, 5, 0], ...            10
4  [[[12, 4, 1], [12, 4, 1], [13, 5, 2], [13, 5, ...             6

   area_contours  len_contours
0       104954.5   5440.028013
1       376831.5   9532.359303
2       342755.5  17617.359080
3       485682.5  12543.058113
4       326303.0  11330.159495
```

xxxxxx

DISEASE	NUMBER OF CONTOURS	AREA OF CONTOURS	LENGTH OF CONTOURS
TINEA CAPITIS	6.32	290119.7	10845.17
ALOPECIA AREATA	4.46	253261.4	8595.91
MELANOMA	6.97	232475.3	9240.73
HEALTHY SCALP	6.67	257113.3	13052.42

7.8 MACHINE LEARNING MODELS

In [59]:
```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.metrics import f1_score, mean_squared_error, accuracy_score
```

In [60]:
```
df=pd.read_csv('/Users/varsha/Desktop/orig2.csv')
df.head()
```

Out[60]:

	ID	Tinea captisis	Alopecia areata	Melanoma	Healthy scalp	Img	pic	segmented_img	colour	num_contours	...	height_width_ratio	central_width	centra
0	1	1	0	0	0	1.jpeg	[[[60 67 75]\n [60 66 75]\n [58 65 ...	[[[0 0 0]\n [0 0 0]\n [0 0 0]\n ...\n [0 0...	(48.514639668367344, 45.396026785714284, 44.45...	3	...	1.515284	229	
1	2	1	0	0	0	2.jpeg	[[[92 81 77]\n [84 73 69]\n [66 55 ...	[[[0 0 2]\n [0 0 2]\n [0 0 2]\n ...\n [0 0...	(51.35983099489796, 42.72837691326531, 35.8404...	6	...	1.235437	412	
2	3	1	0	0	0	3.jpeg	[[[6 58 98]\n [6 58 98]\n [17 69 1...	[[[1 0 0]\n [1 0 0]\n [1 0 0]\n ...\n [0 0...	(36.319072066326534, 28.56017538265306, 30.999...	10	...	0.902128	470	

In [61]:
```
list(df.columns)
```

Out[61]:
```
['ID',
 'Tinea captisis',
 'Alopecia areata',
 'Melanoma',
 'Healthy scalp',
 'Img',
 'pic',
 'segmented_img',
 'colour',
 'num_contours',
 'area_contours',
 'len_contours',
 'area',
 'height_width_ratio',
 'central_width',
 'central_height',
 'smaller_half_dist',
 'circle_area',
 'circle_area_ratio',
 'colour1',
 'zeroes ',
 'skel ',
 'lines']
```

In [62]:
```
df['colour1'] = df['colour1'].map( {'yellow':1, 'pink':2, 'white':3,'peach':4,'red':5,'black':6} )
```

In [63]:
```
X=df[['area','central_height','central_width','height_width_ratio','smaller_half_dist',
 'circle_area','circle_area_ratio','num_contours','area_contours','len_contours','zeroes ','colour1']]
```

In [64]:
```
y = df['Tinea captisis'].to_numpy()
```

In [65]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, shuffle=True)
```

In [66]:
```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
'''df['area']=sc.transform(df['area'])
df['height']=sc.transform(df['height'])
df['width']=sc.transform(df['width'])
df['height_width_ratio']=sc.transform(df['height_width_ratio'])'''

X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
```

In [67]:
```
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(solver='lbfgs')
classifier.fit(X_train, y_train)
```

Out[67]:
```
LogisticRegression()
```

7.8.1 LOGISTIC REGRESSION COMPARISON BETWEEN THE THREE DISEASES AND HEALTHY DATASET:

FOR MELANOMA:

```
Accuracy for melanoma: 0.7794117647058824
[[44  7]
 [ 8  9]]
              precision    recall  f1-score   support

           0       0.85      0.86      0.85        51
           1       0.56      0.53      0.55        17

    accuracy                           0.78        68
   macro avg       0.70      0.70      0.70        68
weighted avg       0.78      0.78      0.78        68
```

FOR TINEA CAPITIS:

```
Accuracy for tinea capitis: 0.7058823529411765
[[41  9]
 [11  7]]
              precision    recall  f1-score   support

           0       0.79      0.82      0.80        50
           1       0.44      0.39      0.41        18

    accuracy                           0.71        68
   macro avg       0.61      0.60      0.61        68
weighted avg       0.70      0.71      0.70        68
```

FOR ALOPECIA AREATA:

```
Accuracy for Alopecia areata: 0.6764705882352942
[[41  5]
 [17  5]]
              precision    recall  f1-score   support

           0       0.71      0.89      0.79        46
           1       0.50      0.23      0.31        22

    accuracy                           0.68        68
   macro avg       0.60      0.56      0.55        68
weighted avg       0.64      0.68      0.63        68
```

FOR HEALTHY SCALP :

```
Accuracy for healthy scalp: 0.8529411764705882
[[54  3]
 [ 7  4]]
              precision    recall  f1-score   support

           0       0.89      0.95      0.92        57
           1       0.57      0.36      0.44        11

    accuracy                           0.85        68
   macro avg       0.73      0.66      0.68        68
weighted avg       0.83      0.85      0.84        68
```

7.8.2 SVM COMPARISON BETWEEN THE THREE DISEASES AND HEALTHY DATASET:

CODE:

```
In [123]: #SVM
from sklearn.svm import SVC
svmclf = SVC()
svmclf.fit(X_train, y_train)

# Storing the predictions of the non-linear model
y_pred = svmclf.predict(X_test)


from sklearn.metrics import r2_score
# Evaluating the performance of the non-linear model
print('Accuracy for healthy scalp : ' +str(accuracy_score(y_test, y_pred)))
from sklearn.metrics import classification_report, confusion_matrix
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
```

FOR HEALTHY SCALP :

```
Accuracy for healthy scalp : 0.8823529411764706
[[57  0]
 [ 8  3]]
              precision    recall  f1-score   support

           0       0.88      1.00      0.93        57
           1       1.00      0.27      0.43        11

    accuracy                           0.88        68
   macro avg       0.94      0.64      0.68        68
weighted avg       0.90      0.88      0.85        68
```

FOR TINEA CAPITIS :

```
Accuracy for tinea capitis : 0.7647058823529411
[[48  4]
 [12  4]]
              precision    recall  f1-score   support

           0       0.80      0.92      0.86        52
           1       0.50      0.25      0.33        16

    accuracy                           0.76        68
   macro avg       0.65      0.59      0.60        68
weighted avg       0.73      0.76      0.73        68
```

FOR ALOPECIA AREATA :

```
Accuracy for Alopecia areata : 0.7794117647058824
[[50  3]
 [12  3]]
              precision    recall  f1-score   support

           0       0.81      0.94      0.87        53
           1       0.50      0.20      0.29        15

    accuracy                           0.78        68
   macro avg       0.65      0.57      0.58        68
weighted avg       0.74      0.78      0.74        68
```

FOR MELANOMA :

```
Accuracy for Melanoma : 0.7647058823529411
[[46  0]
 [16  6]]
              precision    recall  f1-score   support

           0       0.74      1.00      0.85        46
           1       1.00      0.27      0.43        22

    accuracy                           0.76        68
   macro avg       0.87      0.64      0.64        68
weighted avg       0.83      0.76      0.71        68
```

7.8.3 KNN COMPARISON BETWEEN THE THREE DISEASES AND HEALTHY DATASET:

```
In [174]: #KNN
from sklearn.neighbors import KNeighborsClassifier
neigh = KNeighborsClassifier(n_neighbors=3)
neigh.fit(X_train, y_train)
y_pred=neigh.predict(X_test)
print('Accuracy for melanoma: ' +str(accuracy_score(y_test, y_pred)))
from sklearn.metrics import classification_report, confusion_matrix
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
```

FOR MELANOMA :

```
Accuracy for melanoma: 0.7794117647058824
[[45  6]
 [ 9  8]]
              precision    recall  f1-score   support

           0       0.83      0.88      0.86        51
           1       0.57      0.47      0.52        17

    accuracy                           0.78        68
   macro avg       0.70      0.68      0.69        68
weighted avg       0.77      0.78      0.77        68
```

FOR TINEA CAPITIS :

```
Accuracy for Tinea captisis: 0.6764705882352942
[[37 11]
 [11  9]]
              precision    recall  f1-score   support

           0       0.77      0.77      0.77        48
           1       0.45      0.45      0.45        20

    accuracy                           0.68        68
   macro avg       0.61      0.61      0.61        68
weighted avg       0.68      0.68      0.68        68
```

FOR ALOPECIA AREATA :

```
Accuracy for Alopecia areata: 0.6617647058823529
[[38 10]
 [13  7]]
              precision    recall  f1-score   support

           0       0.75      0.79      0.77        48
           1       0.41      0.35      0.38        20

    accuracy                           0.66        68
   macro avg       0.58      0.57      0.57        68
weighted avg       0.65      0.66      0.65        68
```

FOR HEALTHY SCALP :

```
Accuracy for Healthy scalp: 0.8529411764705882
[[55  3]
 [ 7  3]]
              precision    recall  f1-score   support

           0       0.89      0.95      0.92        58
           1       0.50      0.30      0.37        10

    accuracy                           0.85        68
   macro avg       0.69      0.62      0.65        68
weighted avg       0.83      0.85      0.84        68
```

Chapter 8

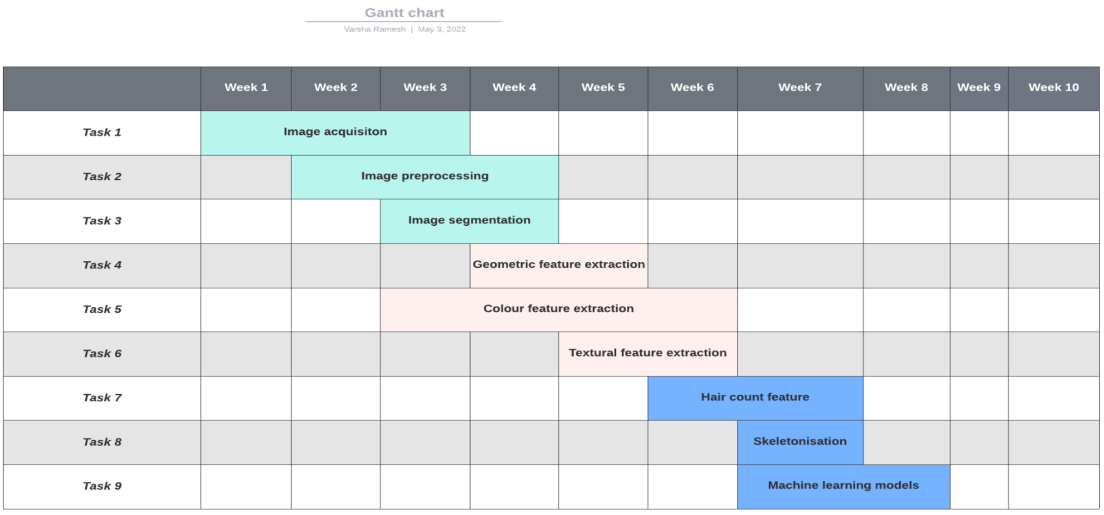
PROJECT PLAN

8.1 INTRODUCTION

The chapter goes through the planning in depth & looks at the various changes that have happened throughout the design and development process.

8.2 GANTT CHART PLAN

The below diagram demonstrates the usage of a Project Charter to divide the different aims and duties. The timetable graphic provided a rapid estimation of the total of effort required to finish a task. Despite the fact that several of the suggestions just on the Network diagram also weren't adopted or performed, the program's main goals were fulfilled.



8.3 PROJECT PLAN OVERVIEW

Due to the obvious machine learning research, the work took awhile to finish. Although being granted a large duration, this program's systems or models infrastructure projects attracted a great deal of attention. The number of hours on Database Design jobs is by far the most significant reason. Machine learning tests and study are a moment and iterative. This is because the algorithms need feature subset modification and a significant chunk of the training process. This was due to machine learning models. A work which was not thoroughly evaluated was indeed the writing of the last annual reports, which took longer than needed to finish, culminating in the deployment not being done correctly.

8.4 CONCLUSION

The part describes the actions that were initially designed or constructed for this program. The Gantt Chart was frequently used to determine durations, and the vast bulk of operations have been finished using this relic. The next part concludes the project.

Chapter 9

CONCLUSION

9.1 INTRODUCTION

The section will go through the model's outputs and consequences, and an assessment of an evidence of concept, upcoming projects to build and extend the health system.

9.2 ALGORITHMS USED

9.2.1 IMAGE PROCESSING ALGORITHMS

Image classification is being used to convert captured pictures to photographic files so that we would execute commands on them and retrieve key characteristics from them. Postprocessing is a type of signal analysis in which the input is a picture, comparable to a camera shot, and the result seems to be the visual or attributes present in an image.

Image enhancement techniques include:

• Visual representation: the ability to perceive quasi things.

special Imaging Search: Identifying an area of interest using image processing techniques.

• Picture repair and deepening: to benefit from this approach to production.

• Patterns linear extrapolation: Examine several items in a picture.

• Object recognition — the ability to distinguish between objects.

9.2.3 MACHINE LEARNING ALGORITHMS

Machine learning design is used in informal community separation, misrepresenting locating, image and speech recognition, audio recognition, PC sight, medical image management, biology, client engagement managers, and a variety of other sectors. ML methods are everywhere, as well as the organizations who are able to prepare neuronal pathways to deliver exceptional results is one of the most sought after professionals.

1. Support vector machine
2. K-NN
3. Logistic regression

9.2.4 MODEL RESULTS

We were able to extract 13 features and inferred multiples values in the tables above, Alopecia Areata has the maximum accuracy using SVM, Melanoma has maximum accuracy using Logistic Regression. Tinea capitis has maximum accuracy using SVM and healthy scalp has high accuracy in all the models as compared to the other diseases.

9.2.5 CHALLENGES

1. Dataset
2. Feature extraction
3. Preprocessing
4. Segmentation

9.2.6 FUTURE WORK

It could be implemented with an app with further details about how it could be treated.

1. **Lee S, Lee JW, Choe SJ, et al. Clinically Applicable Deep Learning Framework for Measurement of the Extent of Hair Loss in Patients With Alopecia Areata. JAMA Dermatol. 2020;156(9):1018–1020. doi:10.1001/jamadermatol.2020.2188**

2. **Ahn CS, Suchonwanit P, Foy CG, Smith P, McMichael AJ. Hair and Scalp Care in African American Women Who Exercise. *JAMA Dermatol.* 2016;152(5):579–580. doi:10.1001/jamadermatol.2016.0093**

3. **W. Chang et al., "A Mobile Device-Based Hairy Scalp Diagnosis System Using Deep Learning Techniques," 2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech), 2020, pp. 145-146, doi: 10.1109/LifeTech48969.2020.1570617332.**

4. **H. Benhabiles et al., "Deep Learning based Detection of Hair Loss Levels from Facial Images," 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2019, pp. 1-6, doi: 10.1109/IPTA.2019.8936122.**

5. **Lacarrubba F, Verzì AE, Micali G. Newly Described Features Resulting From High-Magnification Dermoscopy of Tinea Capitis. *JAMA Dermatol.* 2015;151(3):308–310. doi:10.1001/jamadermatol.2014.3313**

6. **Sunyong Seo, Jinho Park, "Trichoscopy of Alopecia Areata: Hair Loss Feature Extraction and Computation Using Grid Line Selection and Eigenvalue", *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 6908018, 9 pages, 2020. https://doi.org/10.1155/2020/6908018**

7. **Wang, Wei-Chien, Liang-Bi Chen, and Wan-Jung Chang. 2018. "Development and Experimental Evaluation of Machine-Learning Techniques for an Intelligent Hairy Scalp Detection System" Applied Sciences 8, no. 6: 853. https://doi.org/10.3390/app8060853**

8. **Gupta, Aditya, Ivanova, Iordanka, Renaud, Helen, How good is artificial intelligence (AI) at solving hairy problems? A review of AI applications in hair restoration and hair disorders, 34, 10.1111/dth.14811, Dermatologic Therapy**

9. **Development and qualification of a machine learning algorithm for automated hair counting,Jarek P Sacha and Tamara L Caterino and Brian K. Fisher and Gregory J. Carr and Robert Scott Youngquist and Brian D'Alessandro and Anthony Melione and Douglas Canfield and Wilma Fowler Bergfeld and Melissa Peck Piliang and Raghu Kainkaryam and Mike G Davis,International Journal of Cosmetic Science, 2021, 43, S34 - S41**

10. **Sewoong Kim, Jihun Kim, Minjoo Hwang, Manjae Kim, Seong Jin Jo, Minkyu Je, Jae Eun Jang, Dong Hun Lee, and Jae Youn Hwang, "Smartphone-based multispectral imaging and machine-learning based analysis for discrimination between seborrheic dermatitis and psoriasis on the scalp," Biomed. Opt. Express 10, 879-891 (2019)**