INTRODUCTION:

IN TODAY'S WORLD, WHEN HUMAN INTERACTION IS LIMITED AND EARLY DISEASE DETECTION IS CRITICAL, IT IS OF UTMOST IMPORTANCE TO INVEST IN AND CREATE EFFICIENT AND EFFECTIVE TECHNIQUES OF DIAGNOSIS.

THE SCALP IS A SPECIAL SKIN ZONES IN HUMANS, WITH HIGH FOLLICULAR THICKNESS AND A HIGH RATE OF SEBUM CREATION. SEBUM IS UNIQUE IN SOME ASPECTS AND DIFFERS IN COMPOSITION WITH OTHER LIPIDS.

SCALP PROBLEMS HAPPEN IN SOME FACTORS WHICH ARE STRESS, ABUSE OF HAIR PRODUCTS, FOR INSTANCE, STRAIGHTENERS, WAVES, AND DYES AS MAY CAUSE CONTACT ALLERGIC, CHEMICAL BURNS, AND INCREASED HAIR BREAKAGE.

A RELATED ACUTE GENERALIZED DISEASE, PERSONAL PAST HISTORY OF CHRONIC SYSTEMIC DISEASE, AND THE USE OF OTHER DRUGS CAN ALL BE RELATED TO SCALP PROBLEMS

THE MOST COMMON TYPE OF SCALP PROBLEM CONDITIONS EXPERIENCED BY THE HUMAN ARE **ALOPECIA AREATA (AA), ALOPECIA AREATA, MELANONMA. THERE ARE SEVERAL CAUSES INCLUDING NON-MICROBIAL AND MICROBIAL FACTORS.**

IN THIS PROJECT, WE PROPOSE AN ELECTRONIC SCALP EXAMINATION TECHNIQUE DEPENDENT ON QUANTITATIVE AND SUBJECTIVE HIGHLIGHTS LIKE REGION, BALDNESS, WIDTH, SHADING AND SURFACE OF THE SCALP.

PROPOSED SYSTEM:

AS WE ALREADY KNOW THAT HAIR CARE IS EXPENSIVE, THUS IT IS IMPORTANT TO BE ABLE TO DETECT DISEASES EARLY ON, THUS THE PRIMARY OBJECTIVE OF THIS PROJECT IS TO CLASSIFY IMAGES OF SCALP WHICH ARE HEALTHY OR ARE DIAGNOSED WITH EITHER TINEA CAPTISIS, AREATA (AA), ALOPECIA AREATA, MELANONA BY EXTRACTING FEATURES FROM THE SCALP IMAGES

THE SECONDARY OBJECTIVE IS TO PERFORM COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS TO FIND THE BEST PERFORMING MODELS AND THEIR DIFFERENCES IN HOPE TO ACHIEVE BETTER PERFORMANCE THAN THOSE WHICH HAVE BEEN ACHIEVED TILL DATE.

THE THIRD OBJECTIVE IS TO CREATE WEB APP OR MOBILE APP WHERE THE INPUT WOULD BE THE IMAGE AND ALL THE FEATURES WOULD BE EXTRACTED AND WITH THE BEST ACCURACY OF THE MODEL OBTAINED FROM THE SECOND OBJECTIVE WILL BE RUN AND A PREDICTION WILL BE DISPLAYED AS AN OUTPUT AND ADDITIONALLY RESOURCES TO TREAT IT WITH NEAREST HOSPITAL TREATMENTS WILL BE SUGGESTED.

MODULES:

MODULE 2: IMAGE ACQUISITION: THE ACQUISITION PROCESS INVOLVES OBTAINING THE IMAGE BY MAKING USE FROM VARIOUS DATABASES IN VARIOUS WEBSITES. WE ACQUIRED 170 IMAGES AMONG WHICH
47. [TINEA CAPITIS]
46 [ALOPECIA AREATA]
48 [MELANOMA]
29 [HEALTHY SCALP]

MODULE 3: I**MAGE PRE-PROCESSING:** IMAGES ARE RESIZED TO 560X560 FOR COMPUTATIONAL UNIFORMITY AND MODEL CONVERGENCE. A GRAYSCALE CONVERSION IS DONE PRIOR TO

APPROPRIATE FEATURE BEING EXTRACTED. **THE MAIN REASON** WHY GRAYSCALE REPRESENTATIONS ARE OFTEN USED FOR EXTRACTING DESCRIPTORS INSTEAD OF OPERATING ON COLOR IMAGES DIRECTLY IS THAT GRAYSCALE SIMPLIFIES THE ALGORITHM AND REDUCES COMPUTATIONAL REQUIREMENTS

MODULE 4: NOISE REMOVAL : A FIVE BY FIVE KERNEL IS USED AS A 2D FILTER FOR NOISE REMOVAL. KERNEL FILTER IS A TYPE OF KERNEL-BASED SMOOTHING METHOD WHICH IS OFTEN APPLIED TO AN INPUT VECTOR, TIME SERIES OR MATRIX TO GENERATE THE SMOOTHED VERSION OF THE INPUT SEQUENCE. **THE MAIN REASON :** IMAGE PROCESSING FILTERS ARE MAINLY USED TO SUPPRESS EITHER THE HIGH FREQUENCIES IN THE IMAGE, I.E. SMOOTHING THE IMAGE, OR THE LOW FREQUENCIES, I.E. ENHANCING OR DETECTING EDGES IN THE IMAGE. THE FILTER FUNCTION IS SHAPED SO AS TO ATTENUATE SOME FREQUENCIES AND ENHANCE OTHERS.

MODULE 5: IMAGE SEGMENTATION: TO SEGMENT THE IMAGES FOREGROUND-BACKGROUND EXTRACTION USING THE GRABCUT ALGORITHM. THE GAUSSIAN MIXTURE MODEL IS USED FOR MODELLING THE FOREGROUND AND BACKGROUND BY WHICH THE PIXELS IN THE REGION OF INTEREST ARE LABELLED AND CLUSTERED.
THE MAIN REASON : THE GRABCUT ALGORITHM IS CHOSEN OVER OTHER METHODS DUE TO THE NON-UNIFORMITY OF THE BACKGROUND. THIS ALGORITHM PERFORMS WELL IN SUCH A CONDITION AND THE REASON LAYS AT UNCERTAINTY PRINCIPLE WHERE POSITIONS AND FREQUENCIES OF THE SIGNAL CANNOT BE DETERMINED AT THE SAME TIME.

MODULE 6: FEATURE EXTRACTION: EXTRACTING FEATURES SUCH AS SCALP BOTCH (BALDING), THICKNESS OF THE HAIR, TEXTURE, CHROMATIC AND GEOMETRIC.

MODULE 7: WHY THE RESPECTIVE ML MODELS?

THE FEATURES EXTRACTED FROM THE IMAGES ARE USED TO BUILD THREE DIFFERENT CLASSIFICATION MODELS, SUPPORT VECTOR CLASSIFIER (SVC), LOGISTIC REGRESSION MODEL AND K-NN.

SUPPORT VECTOR MACHINE ARE LEARNING MODELS WHICH IS USED FOR CLASSIFICATION AND REGRESSION ANALYSIS. IT IS SPECIFICALLY USED FOR SUPERVISED BINARY CLASSIFICATION. IT USES THE CONCEPT OF KERNEL TRICK TO FIND AN OPTIMAL BOUNDARY BETWEEN THE OUTPUTS.

LOGISTIC REGRESSION MAINLY FOCUSES ON CLASSIFICATION. IT IS GENERALLY USED FOR THE PREDICTION OF BINARY OUTPUTS. IT ANALYSES THE RELATIONSHIP BETWEEN A DEPENDENT AND AN INDEPENDENT VARIABLE.

KNN ALGORITHM IS WHEN A SEARCH IS PERFORMED ON THE TRAINING DATASET TO GET THE CLOSEST LABEL FOR THE DATA PRESENT IN THE DATASET. WE ARE CHOOSING THE NEAREST NEIGHBOURS AND FIND SIMILAR ATTRIBUTES WHICH HELP US IN CLASSIFICATION.

MODULE 8: WEB/MOBILE APPLICATION THAT CAN TAKE INPUT AS AN IMAGE AND THE OUTPUT WILL BE THE CLASSIFICATION OF THE DISEASE

FEATURE EXTRACTION:

FOR CHROMATIC FEATURE:
RGB VALUES FOR 5 COMMONLY OBSERVED COLOURS IN TONGUES ARE DEFINED AS REFERENCE VALUES TO WHICH THE PIXELS IN THE IMAGE WILL BE COMPARED. THE MANHATTAN DISTANCE,AS IN , IS CALCULATED BETWEEN THE RGB VALUES OF THE PIXELS AND THE BASE COLOUR VALUES. THE COLOUR OF AN IMAGE IS THEREFORE IDENTIFIED AS THE REFERENCE COLOUR TO WHICH A MAJORITY OF THE FOREGROUND PIXELS HAVE A MINIMUM DISTANCE.

TO FIND THE HAIR COUNT WE MAKE USE OF THE LINE DETECTION ALGORITHM.

WHAT IS THE LINE DETECTION ALGORITHM?
IN IMAGE PROCESSING, LINE DETECTION IS AN ALGORITHM THAT TAKES A COLLECTION OF 'N' EDGE POINTS AND FINDS ALL THE LINES ON WHICH THESE EDGE POINTS LIE.

THE MOST POPULAR LINE DETECTORS ARE THE HOUGH TRANSFORM AND CONVOLUTION-BASED TECHNIQUES.

WHY I CHOSE HOUGH TRANSFORM OVER CONVOLUTION BASED TECHNIQUES?

THE ADVANTAGE OF THE HOUGH TRANSFORM OVER CONVOLUTION BASED TECHNIQUE IS THAT PIXELS LYING ON ONE LINE NEED NOT ALL BE CONTIGUOUS. THIS CAN BE VERY USEFUL WHEN TRYING TO DETECT LINES WITH SHORT BREAKS IN THEM DUE TO NOISE, OR WHEN OBJECTS ARE PARTIALLY OCCLUDED.

WHAT IS HOUGH TRANSFORM?

HOUGH TRANSFORM IS A POPULAR TECHNIQUE TO DETECT ANY SHAPE, IF YOU CAN REPRESENT THAT SHAPE IN MATHEMATICAL FORM. IT CAN DETECT THE SHAPE EVEN IF IT IS BROKEN OR DISTORTED A LITTLE BIT.
THE HOUGH TRANSFORM IS A FEATURE EXTRACTION TECHNIQUE USED IN IMAGE ANALYSIS, COMPUTER VISION, AND DIGITAL IMAGE PROCESSING. THE PURPOSE OF THE TECHNIQUE IS TO FIND IMPERFECT INSTANCES OF OBJECTS WITHIN A CERTAIN CLASS OF SHAPES BY A VOTING PROCEDURE.

HOW IT WORKS?

ANY LINE CAN BE REPRESENTED IN THESE TWO TERMS, (ρ, θ). SO FIRST IT CREATES A 2D ARRAY OR ACCUMULATOR (TO HOLD VALUES OF TWO PARAMETERS) AND IT IS SET TO 0 INITIALLY. LET ROWS DENOTE THE ρ AND COLUMNS DENOTE THE θ. SIZE OF ARRAY DEPENDS ON THE ACCURACY YOU NEED.

IT SIMPLY RETURNS AN ARRAY OF (ρ, θ) VALUES. ρ IS MEASURED IN PIXELS AND θ IS MEASURED IN RADIANS.

FIRST PARAMETER, INPUT IMAGE SHOULD BE A BINARY IMAGE, SO APPLY THRESHOLD OR USE CANNY EDGE DETECTION BEFORE FINDING APPLYING HOUGH TRANSFORM.

SECOND AND THIRD PARAMETERS ARE ρ AND θ ACCURACIES RESPECTIVELY.

FOURTH ARGUMENT IS THE THRESHOLD, WHICH MEANS MINIMUM VOTE IT SHOULD GET FOR IT TO BE CONSIDERED AS A LINE.

REMEMBER, NUMBER OF VOTES DEPEND UPON NUMBER OF POINTS ON THE LINE.

SO IT REPRESENTS THE MINIMUM LENGTH OF LINE THAT SHOULD BE DETECTED.

BALDING OF THE HAIR:
We make use of skeletonisation. Skeletonization is a process for reducing foreground regions in a binary image to a skeletal remnant that largely preserves the extent and connectivity of the original region while throwing away most of the original foreground pixels.
It preserves the topology (retains the topology of the original object), It preserves the shape (significant feature suitable for object recognition or classification is extracted), It forces the "skeleton" being in the middle of the object, and.

GEOMETRIC FEATURES

VARIOUS GEOMETRIC FEATURES, INCLUDING DIMENSIONS AND THEIR RATIOS, ARE EXTRACTED FROM THE IMAGES IN ADDITION TO THE CHROMATIC AND TEXTURAL FEATURES.

5.3.1 AREA: AREA IS CALCULATED BY MEASURING THE NUMBER OF COLOURED PIXELS.
5.3.2. HEIGHT-WIDTH RATIO(HW): IT IS THE RATIO OF THE HEIGHT OF THE SCALP (H) AND THE WIDTH OF THE TONGUE(W) AS GIVEN IN (2).
5.3.3. CENTRAL WIDTH: THE WIDTH IS MEASURED AS THE HORIZONTAL DISTANCE FROM THE CENTRAL PIXEL TO THE FARTHEST POINT ON THE X AXIS.
5.3.4. CENTRAL HEIGHT: THE HEIGHT IS MEASURED AS THE VERTICAL DISTANCE FROM THE CENTRAL PIXEL TO THE FARTHEST POINT ON THE Y AXIS.
5.3.5. SMALLER HALF DISTANCE: SMALLER HALF DISTANCE IS THE HALF DISTANCE OF EITHER THE HEIGHT OR THE WIDTH, THE CONDITION FOR IT TO BE CHOSEN IS DEPENDING ON WHICH OF THEM IS SHORTER.
5.3.6. CIRCLE AREA: IT IS DEFINED AS THE AREA OF THE CIRCLE PRESENT ON THE PALETTE OF THE TONGUE USING THE FORMULA OF THE SMALLER HALF DISTANCE.
5.3.7. CIRCLE AREA RATIO: IT IS THE RATIO BETWEEN THE CIRCLE AREA AND THE AREA OF THE SCALP

TEXTURAL FEATURES:

MEANINGFUL TEXTURE IS EXTRACTED FROM THE RAW IMAGES WHICH IS USED TO IDENTIFY COATING, SMOOTHNESS, LINES AND SMALL PORES EACH REFLECTING DIFFERENT PATHOLOGICAL CHANGES IN THE BODY.

SEGMENTATION CODE:

I HAVE TAKEN INPUT AS AN IMAGE AND APPLIED A MASK ON IT, ALONG WITH THAT, I HAVE DECLARED MAY BACKGROUND MODEL AND FOREGROUND MODEL AND I HAVE ASSIGNED VALUES FOR MY RECTANGLE WITH THE STARTING X COORDINATE, Y COORDINATE AND THE WIDTH AND THE HEIGHT

THEN, I HAVE APPLIED THE GRAB CUT ALGORITHM ALONG WITH THE MASK

COLOUR FEATURE CODE:

I HAVE ASSIGNED MY OWN RGB VALUES INSTEAD OF DEFAULT ONES.

THEN USING THE MANHATTAN DISTANCE, WHERE I HAVE DISCUSSED THE FORMULA EARLIER, WE CALCULATE THE MANHATTAN DISTANCE AND RETURN THE COLOUR

WE CALCULATE THE PIXEL COUNT BY FINDING THE IMAGE WIDTH AND IMAGE HEIGHT AND BASED ON THE PIXEL RANGE, WE ASSIGN A VALUE

GEOMETRIC FEATURE CODE:

5.3.1 AREA: AREA IS CALCULATED BY MEASURING THE NUMBER OF COLOURED PIXELS.

WE APPLY A 2D FILER AND START FROM THE I AND J COORDINATE TO CALCULATE THE PIXELS

5.3.2. HEIGHT-WIDTH RATIO(HW): IT IS THE RATIO OF THE HEIGHT OF THE SCALP (H) AND THE WIDTH OF THE TONGUE(W) AS GIVEN IN (2).

5.3.3. CENTRAL WIDTH: THE WIDTH IS MEASURED AS THE HORIZONTAL DISTANCE FROM THE CENTRAL PIXEL TO THE FARTHEST POINT ON THE X AXIS.
5.3.4. CENTRAL HEIGHT: THE HEIGHT IS MEASURED AS THE VERTICAL DISTANCE FROM THE CENTRAL PIXEL TO THE FARTHEST POINT ON THE Y AXIS.

WE ASSIGN A MIDDLE X AND Y COORDINATE AS 200 AND WE INITIALISE TOP, LEFT, BOTTOM AND RIGHT AS 0 AND THEN WE RUN THE I LOOP UNTIL OUR MID X COORDINATE AND INCREASE THE LEFT VALUE BY 1

SIMILARLY FOR TOP,BOTTOM, RIGHT

THEN TO CALCULATE CENTRAL WIDTH, WE ADD LEFT AND RIGHT

FOR CENTRAL HEIGHT, WE ADD TOP AND BOTTOM

THEN FOR HEIGHT WIDTH RATIO, WE DIVIDE THE HEIGHT AND THE WIDTH

5.3.5. SMALLER HALF DISTANCE: SMALLER HALF DISTANCE IS THE HALF DISTANCE OF EITHER THE HEIGHT OR THE WIDTH, THE CONDITION FOR IT TO BE CHOSEN IS DEPENDING ON WHICH OF THEM IS SHORTER.
5.3.6. CIRCLE AREA: IT IS DEFINED AS THE AREA OF THE CIRCLE PRESENT ON THE PALETTE OF THE TONGUE USING THE FORMULA OF THE SMALLER HALF DISTANCE.
5.3.7. CIRCLE AREA RATIO: IT IS THE RATIO BETWEEN THE CIRCLE AREA AND THE AREA OF THE SCALP

SKELETONISATION CODE:

WE GET THE SHAPE OF THE IMAGE USING NUMPY.ZEROS AND WE RETURN THE IMAGE BY APPLYING A THRESHOLD ON THE SHAPE OF THE IMAGE, WITH THAT IMAGE, WE GET THE STRUCTURING ELEMENT USING A MORPH CROSS. WITH THIS IMAGE WE DILATE, ERODE AND SUBTRACT THE IMAGE. WE THEN PERFORM BITWISE OR ON THE IMAGE TO GET THE SKELETON.

HAIR COUNT CODE:

THE GRAY IMAGE IS CONVERTED TO A THRESHOLD IMAGE, THEN WITH AN INBUILT IMAGE, WE FIND THE CONTOURS USING CV2.RETR_EXTERNAL, CV2.CHAIN_APPROX_SIMPLE AND THEN WE THEN WRITE A CONDITION TO DRAW CONTOURS ON THE IMAGE, AND INCREMENT IT TO KEEP A COUNT OF THE LINES.

TEXTURAL FEATURE CODE:

NUMBER OF CONTOURS: THE NUMBER OF CONTOURS ALONG WITH THE AREA OF CONTOURS AND THE PER CONTOUR AREA WAS ACHIEVED BY CONVERTING THE GRAY SCALE IMAGE BY USING A GAUSSIAN BLUR FILTER AND BY APPLYING AN ADAPTIVE THRESHOLD ON THE BLURRED IMAGES. NEXT, THE CONTOURS WERE CALCULATED USING THE THRESHOLDED IMAGES AND A CONDITION WAS GIVEN TO FILTER THE IMAGES THAT HAD THE TEXTURAL IMAGES.

AREA OF THE CONTOUR: ONCE WE GET THE NUMBER OF CONTOURS, WE CHECK IF THE NUMBER OF CONTOURS IS LESS THAN 1000, AND APPEND IT TO THE AREA

LENGTH OF THE CONTOUR : ONCE WE GET THE AREA OF THE CONTOUR, WE CALCULATE THE LENGTH OF THE CONTOUR BY USING CV2.ARCLENGTH() AND APPEND IT BY ADDING THE TOTAL TO THE LENGTH OF THE CONTOUR.

Type of Scalp Condition	Feature properties		Range Value
Alopecia Areata (AA)	Color	Mean Red	129.566-200.247
		Mean Green	82.383-139.862
		Mean Blue	60.721-130.12
	Texture	Contrast	24.3587-3396.99
		Correlation	0.607236-0.997313
		Energy	0.000104585-9.22461
		Homogeneity	0.0895999-0.553896
	Shape	Area	1.12562-386284
		Major Axis	144.338-3491.81
		Minor Axis	101.614-2327.88
Dandruff	Color	Mean Red	133.135-169.959
		Mean Green	109.333-133.872
		Mean Blue	0.71-131.301
	Texture	Contrast	4.4669-3587.24
		Correlation	0.6007-0.933709
		Energy	0.000108172-8.50067
		Homogeneity	0.0691-0.207285
	Shape	Area	1.31808-1177056
		Major Axis	185.907-287280
		Minor Axis	166.277-68904
Normal	Color	Mean Red	125.697-188.517
		Mean Green	89.863-142.996
		Mean Blue	64.442-135.006
	Texture	Contrast	55.8271-3144.61
		Correlation	0.689855-0.994537
		Energy	0.000115807-9.7932
		Homogeneity	0.0962138-0.457836
	Shape	Area	1.87488-383671
		Major Axis	155.885-1939.9
		Minor Axis	125.862-1288.65