



The Role of Artificial Intelligence in Curly Hair Analysis

Prepared By

Name: SARA ISMAIL

ID No [252000512](#)

Submitted in fulfillment of the requirements for the degree Master of Science in Artificial Intelligence

At

Collage of Informatics

Midocean University

UAE

Supervisor Name: Dr Khaled Eskaf

Collage of Informatics, Midocean University, Comoros

Email: saraismailse@gmail.com

Acknowledgements

Special thanks to Dr. Khaled Eskaf and Dr. Sherif Mostafa for their continuous guidance, valuable feedback, and unwavering support throughout the development of this thesis.

Abstract

Curly hair analysis is a challenging task due to the variability in hair types and the difficulty of assessing hair damage using traditional methods. This thesis proposes an artificial intelligence-based approach for analyzing curly hair by combining deep learning and machine learning techniques.

First, a deep learning model based on EfficientNet-B0 was used to classify different hair types from regular images, including straight, wavy, curly, and kinky hair. Several experiments were performed by tuning different parameters to improve model performance.

Second, scanning electron microscope (SEM) image analysis was conducted to evaluate hair surface conditions. Feature extraction techniques were applied to obtain relevant characteristics, followed by clustering methods to classify hair damage levels.

The experimental results demonstrated that the proposed approach achieved promising performance in both hair type classification and hair damage analysis. Finally, the outputs of both models were used to support the development of a future recommendation system based on hair type and damage level.

The proposed framework provides a step toward building intelligent systems for personalized hair care

Table of Contents

Acknowledgements.....	3
Abstract.....	4
List of Abbreviations.....	8
List of Figures.....	9
List of Tables.....	9
Chapter 1: Introduction.....	1
1.1 Problem Definition.....	1
1.2 Thesis Objectives.....	1
1.3 Thesis Contribution.....	2
1.4 Thesis Organization.....	2
Chapter 2: Background and Related Work.....	3
Background.....	3
2.1 Introduction.....	3
2.2 Artificial Intelligence (AI)	3
2.3.2. Machine Learning (ML).....	3
2.4 Deep Learning (DL).....	4
2.5 convolutional neural networks (CNN).....	4
2.6 Scanning Electron Microscopy (SEM).....	5
2.7 Region of Interest (ROI).....	5
2.8 Tailing Factor.....	5
2.9 K-means clustering	6
2.10 EfficientNet-B0.....	6
Related Work.....	7
Chapter 3 Proposed Approach (Methodology).....	15
3.1 introduction.....	15
3.2 Dataset Description.....	15
3.2.1 Hair Type Classification.....	15
3.2.2 SEM Images.....	16
3.3 Data Preprocessing.....	16
3.3.1 Hair Type Images Preparation.....	16
3.3.2 SEM Image Preparation.....	17
3.4 Deep Learning Model.....	17
3.4.1 Training Process.....	17

3.4.2 Classification Task.....	17
3.5 Machine Learning Model for SEM Analysis.....	18
3.5.2 Feature Extraction.....	18
3.6 Proposed Framework.....	18
3.7 Evaluation Metrics.....	19
3.8 Summary.....	19
Chapter 4 :Experimental Results and Discussion.....	19
4.1 Introduction.....	20
4.2 Curly Hair Type Classification.....	20
4.2.1 Experiment 1.....	20
4.2.2 Experiment 2.....	22
4.2.3 Experiment 3.....	23
4.2.4 Experiment 4.....	24
4.2.5 Experiment Summary.....	25
4.3 SEM Analysis.....	26
4.3.1 Dryad Dataset Experiments.....	27
Experiment 1 - RIO 75% - FILTER_MAX_4000X FALSE.....	27
Experiment 2 - RIO 70% - FILTER_MAX_4000X FALSE.....	28
Experiment 3 - RIO 50% - FILTER_MAX_4000X FALSE.....	29
Experiment 4 - RIO 40% - FILTER_MAX_4000X FALSE.....	30
Experiment 5 - RIO 40% FILTER_MAX_4000X TRUE.....	31
4.3.2 L'Oréal Dataset Experiment.....	32
Chapter 5 Conclusion and Future Work.....	34
5.1 Conclusion.....	34
5.2 Future Work.....	34
References.....	35
Datasets.....	37

List of Abbreviations

- **CH:** Curly Hair
- **AI:** Artificial Intelligence
- **DL:** Deep Learning
- **ML:** Machine Learning
- **CV:** Computer Vision
- **CNNs:** Convolutional Neural Networks
- **SEM:** Scanning Electron Microscopy
- **RIO:** Region of Interest

List of Figures

Fig. 1. AI and its subfields	3
Fig. 2. Deep Learning	4
Fig. 3. CNNs	4
Fig. 4. SEM Image	5
Fig. 6. Tailing Factor	5
Fig. 7. K-means Clustering	16
Fig.8. Hair type classification	18
Fig. 9. SEM analysis	19
Fig. 10. Hair Type Classification - Experiment 1	20
Fig. 11. Hair Type Classification - Experiment 1 – confusion matrix	21
Fig. 12. Hair Type Classification - Experiment 2	22
Fig. 13. Hair Type Classification - Experiment 2 – confusion matrix	22
Fig. 14. Hair Type Classification - Experiment 3	23
Fig. 15. Hair Type Classification - Experiment 3 – confusion matrix	23
Fig. 16. Hair Type Classification - Experiment 4	24
Fig. 17. Hair Type Classification - Experiment 4 – confusion matrix	24
Fig. 18. SEM Analysis - Experiment 1- RIO	27
Fig. 19. SEM Analysis - Experiment 1 – Tailing factor distribution	27
Fig. 20. SEM Analysis - Experiment 2- RIO	28
Fig. 21. SEM Analysis - Experiment 2 – Tailing factor distribution	28
Fig. 22. SEM Analysis - Experiment 3- RIO	29
Fig. 23. SEM Analysis - Experiment 3 – Tailing factor distribution	29
Fig. 24. SEM Analysis - Experiment 4- RIO	30
Fig. 25. SEM Analysis - Experiment 4 – Tailing factor distribution	30
Fig. 26. SEM Analysis - Experiment 5- RIO	31
Fig. 27. SEM Analysis - Experiment 5 – Tailing factor distribution	31
Fig. 28. SEM Analysis – L'Oréal RIO	32
Fig. 29. SEM Analysis – L'Oréal Tailing factor	32
Fig. 30. SEM Analysis – Kmeans clustering by Tailing factor	33

List of Tables

Table 1 – hair type dataset	15
Table 2 – SEM datasets	16
Table 3 – Hair Type Classification - Experiment Summary	25

Chapter 1: Introduction

Curly Hair (CH) has received attention in recent years due to the growing interest in hair health care. However, analyzing and classifying CH types accurately remains challenging due to the wide variety of CH textures, in addition to identifying hair damage is still difficult, especially when depending on manual assessment.

The field of artificial intelligence (AI), particularly Machine Learning (ML), Deep Learning (DL) and Computer Vision (CV) have recently experienced rapid development and we see its strong presence in many fields, including healthcare. These technologies enable systems to learn complex patterns and improve prediction accuracy in various domains; this has highlighted the need to use AI in CH analysis.

Therefore, this research explores the role of AI in analyzing CH and proposes an AI model to classify CH types and analyze hair damage using image-based techniques to provide accurate hair analysis results.

1.1 Problem Definition

Classifying and analyzing CH is challenging due to the wide variety of CH textures, and classification is often done manually based on personal experience, and measuring hair damage also requires accurate analysis, and prior knowledge for microscopic images like Scanning Electron Microscopy (SEM) images, which are difficult to analyze.

Also, there are currently few tools that can classify hair type and analyze hair damage. The AI and DL techniques have made it possible to analyze images more accurately and extract valuable features, but there is still an insufficient system specifically designed for analyzing CH.

Therefore, the difficulty of classifying CH types and analyzing hair damage accurately is the main challenge in this research.

1.2 Thesis Objectives

The core objective of this thesis is to design and evaluate intelligent systems capable of classifying and analyzing CH automatically based on images and analyze SEM images to define hair damage degree by integrating feature extraction methods to deliver reliable predictions. In conclusion, the goal is to provide a model to predict the type of hair and the degree of damage accurately.

1.3 Thesis Contribution

This thesis has the following main contributions in the field of CH analysis:

- 1- develop an AI based model to classify CH type based on image.
- 2- propose a model to analyze SEM images to define hair damage.
- 3- Using a DL model to classify CH types using images relies on Convolutional Neural Networks (CNNs), the EfficientNet model will be used as the core framework.
- 4- Using ML basic learners including KMeans clustering to find patterns in SEM images to make predictions more reliable and applicable to a wider range of situations.

1.4 Thesis Organization

The remainder of the thesis is structured as follows:

Chapter 2 Introduces core AI concepts such as ML, DL and CV and explains their roles in image analytics. It will explain the SEM images and discuss the importance of feature extraction. The chapter will examine recent academic research and studies related to image analysis, especially hair images and SEM images. It identifies current research gaps that this thesis aims to address.

Chapter 3 Describes the research design and methodology, including data collection, preprocessing, algorithm selection, and the integration of feature extraction technique. It explains how different classification models and feature extraction methods are applied and evaluated.

Chapter 4 Reports the experimental setup and presents results obtained from various classification models. The chapter includes a detailed performance analysis. It also discusses the feature extraction method.

Chapter 5 Summarizes the key findings, contributions, and challenges of the study. It outlines potential directions for future work in the AI recommendation system for CH analysis.

Chapter 2: Background and Related Work

Background

2.1 Introduction

AI and its subfields contribute to hair analysis. ML, DL, and CV each concept is illustrated to show how it supports image analysis. CV and DL enable us to build predictive models to predict hair type, and with ML, we can analyze SEM images and extract features like tilling factors to predict the damage in hair cuticles.

In this chapter, we will focus on the key principles of the CH analysis using AI, and how AI and its subfields can be applied in this field. Also, we will discuss feature extraction techniques and key algorithms for classification, as well as the importance of extracted features.

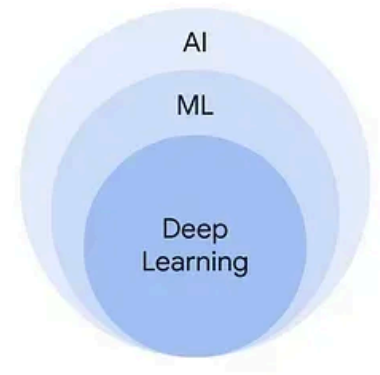


Fig. 1. AI and its subfields

2.2 Artificial Intelligence (AI)

AI is a set of technologies for simulating human intelligence, through a set of approaches to perform tasks that require human intelligence, like learning.

Systems are learned from huge amounts of data; they analyze data and identify patterns to make predictions or decisions and provide helpful suggestions without being explicitly programmed.[1]

2.32. Machine Learning (ML)

ML is a subset of AI that focuses on developing algorithms to learn from data, identify patterns, and make predictions or decisions based on its learnings. It applies different techniques to learn such as supervised learning, unsupervised learning, and reinforcement learning [1]

In CH analysis ML used to analyze SEM images and predict the degree of damage.

2.4 Deep Learning (DL)

DL is a subfield of ML that uses neural networks to simulate the complex decision-making of the human brain. In CH analysis DL used for image recognition and classification to identify hair pattern types.

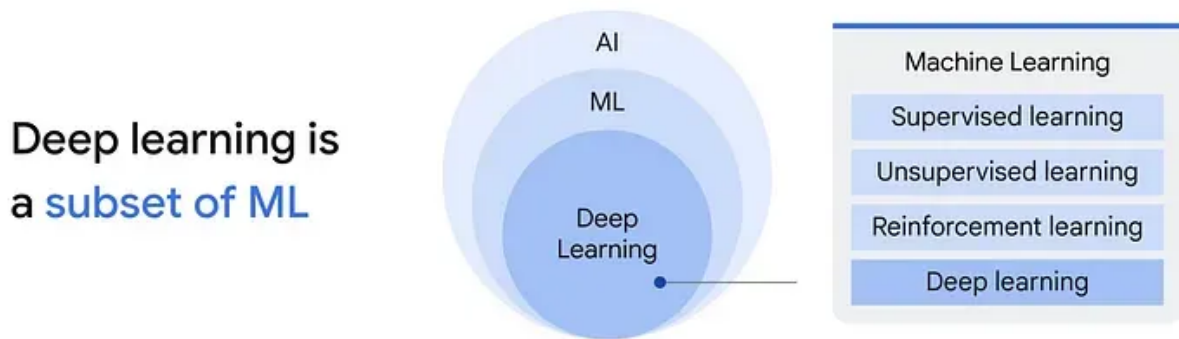


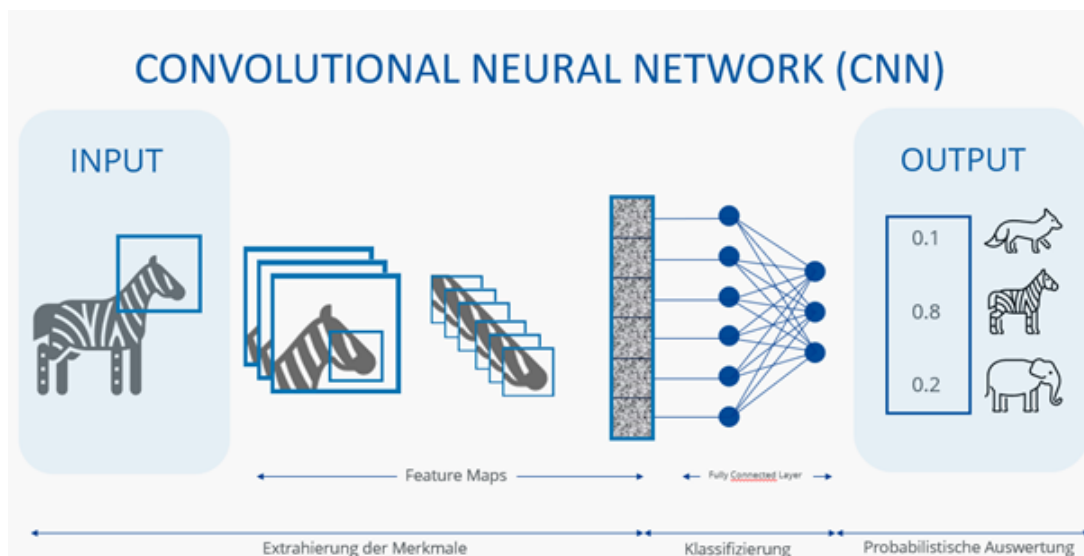
Fig. 2. Deep Learning

2.5 convolutional neural networks (CNN)

CNNs use three-dimensional data for image classification and object recognition; they are at the heart of deep learning algorithms.

CNNs are more often utilized for classification and computer vision tasks. It provides a scalable approach for image classification and object recognition tasks, and can be computationally demanding, requiring graphical processing units (GPUs) to train models. [2]

Fig. 3. CNNs



2.6 Scanning Electron Microscopy (SEM)

SEM is a scientific instrument that uses a focused beam of electrons to examine the detailed surface characteristics and three-dimensional structure of a specimen, providing high-resolution images. It bridges the gap between the magnification capabilities of light microscopy and the higher resolving power of the transmission electron microscope.[3]

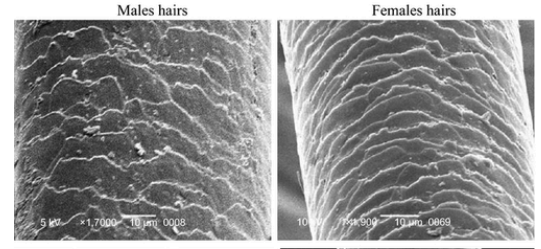


Fig. 4. SEM Image

SEM images are used to extract feature (tailing factor) to identify the degree of hair damage

2.7 Region of Interest (ROI)

ROI is a selected portion of an image that is isolated for more analysis. It allows researchers to focus on relevant areas while excluding irrelevant areas.

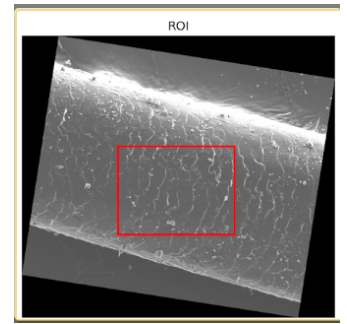


Fig. 5. ROI from SEM image

2.8 Tailing Factor

The tailing factor is a measure of asymmetry in pixel brightness histograms and proxy for surface roughness. It serves as a proxy for hair surface roughness and damage severity. [4]

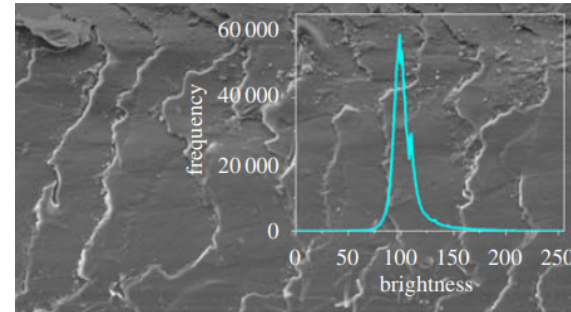


Fig. 6. Tailing Factor

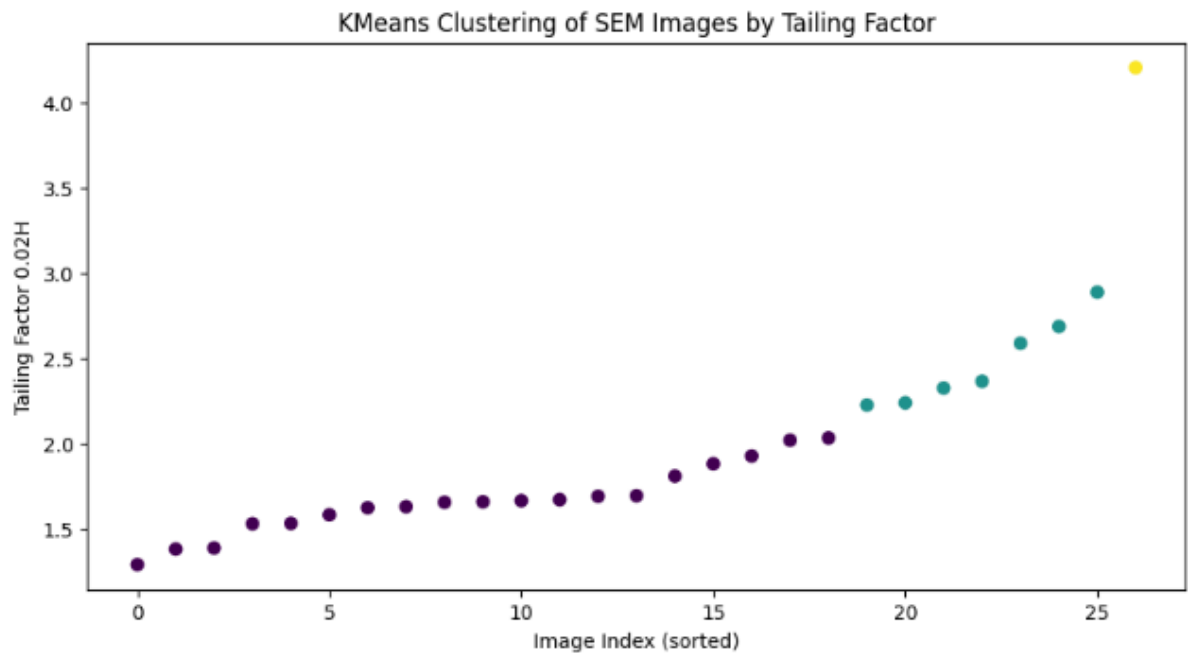
In this research ROI has been chosen from SEM to exclude hair edges and areas with unnatural lighting, then a tailing factor calculated as a proxy for hair surface roughness and damage severity.

2.9 K-means clustering

K-means clustering is an unsupervised learning algorithm used for data clustering, which groups unlabeled data points into groups or clusters.[5] It is efficient, effective and simple clustering method used for clustering SEM images based on tailing factor for 3 clusters

- 1- Low damage
- 2- Medium damage
- 3- High damage

Fig. 7. K-means Clustering



2.10 EfficientNet-B0

EfficientNet is a highly optimized convolutional neural network (CNN) designed for environments with limited computational resources designed for image classification.

EfficientNet-B0 is used to classify hair type from 4 classes:

- 1- Curly
- 2- Kinky
- 3- Straight
- 4- Wavy

Related Work

Fanny Chu [4] proposed a Quantitative technique to measure the degree of the damage

An automated method developed to analyze SEM images and measure hair damage without relying on manual assessment. The research discusses that hair damage appears in SEM images such as cracks, holes, cuticle damage and rough surface, therefore, the damaged areas show more brightness, so they used a Histogram brightness for analysis

To measure hair damage, the research presented the Tailing Factor as a method for measuring damage (the tailing factor is a measure of histogram asymmetry) and used as indicator for hair roughness and hair damage, the steps they followed were:

Use a SEM image as an input

1. Determine ROI
2. Normalize Image
3. Histogram extraction
4. Brightness calculation
5. Tailing factor calculation
6. Hair classification by degree of damage

In the research the ROI was chosen to exclude hair edges and areas with unnatural lighting and analyze only the important part of the hair. The results show that the tailing factor succeeded in classifying damaged hair, and accuracy reached 81%.

While this research focused on quantifying damage, this study did not address hair type classification or provide personalized product recommendations. This research extends previous work by integrating artificial intelligence techniques to classify hair type, assess hair damage, and recommend suitable hair care products (in the next stage).

Muhammad UR, [6] discusses the factors that make hair analysis difficult due to the variety of hair patterns, variety of hair colors, complex image background and most of previous analysis methods need an image from front view. The research proposed a method to analyze hair by detecting hair in image and do hair segmentation then classify hair type pattern.

The research was based on texture analysis, computer Vision and deep Learning, a CNN features, random Forest and SVM have been used, and different hair patterns and hair styles have been classified into wavy, curly, kinky, braids, dreadlocks and Short

The research proposed new dataset (Figaro1k) with more than 1000 images, ready segmentation masks and 7 different types of hair images, and accuracy reached 90%

While this research focused on hair type classification, this study did not address hair damage analysis or provide personalized product recommendations. This research extends previous work by integrating artificial intelligence techniques to classify hair type, assess hair damage, and recommend suitable hair care products (in the next stage).

Dertli B, Koklu M. [7] A research aimed to classify hair types using DL, by using a CNN model to extract features from images then a ML classifier for final classification.

The research stated that most studies focused on hair color and hair loss, while there is no robust study classifying structural hair types such as straight, wavy, curly, kinky and dreadlocks. A Hair Type Dataset with 1992 image has been used, and an image was resized to 224×224. Three Deep Learning models were used to extract features (SqueezeNet, InceptionV3 and VGG19)

And a classification algorithm used to classify hair types (SVM, ANN, KNN, and random forest) and the performance of each model was compared.

The best results were:

- 1- InceptionV3 + ANN: accuracy = 84.9%
- 2- SqueezeNet + ANN: accuracy = 84.4%
- 3- VGG19 + ANN: accuracy = 79.7%

Important observations from the research InceptionV3 provided the best texture representation; VGG19 was accurate but required higher computational power; SqueezeNet was fast and lightweight.

While this research focused on hair type classification, this study did not address hair damage analysis or provide personalized product recommendations. This research extends previous work by integrating artificial intelligence techniques to classify hair type, assess hair damage, and recommend suitable hair care products (in the next stage).

Also the research relied on models such as VGG19, InceptionV3, and SqueezeNet while this research utilizes EfficientNet, which is known for its improved performance and computational efficiency. EfficientNet achieves higher accuracy with fewer parameters, making it suitable for hair classification tasks.

Daniels, G., Tamburic, S. [8], presents the use of artificial intelligence (AI) to evaluate the effect of hair care products on hair volume, straightness, and flyaways, this done by analyzing images of straight hair treated with different products, using machine learning algorithms, 2D image analysis, and human tests for comparison. It explains why hair shape and appearance are important and demonstrates that consumers are interested in modifying hair to suit their taste. It also mentions that companies produce variant hair care products to determine their effectiveness and they conduct evaluation tests often performed by manual human visual assessments, which are sometimes inaccurate.

This study has two main objectives, first one is to use computer vision and machine learning techniques to measure the effect of treatments on hair assembly, so that it is possible to

identify whether hair volume has increased or its appearance has changed after a particular treatment, and the second one is to compare the performance of AI with human assessments. It also demonstrates that shape, size, and flyaway are not just aesthetic, but also result from physical and chemical properties, thus reinforcing the importance of automated analysis tools such as AI to measure these properties more precisely.

Hill K, Adesola A, Dangal P, [21] proposes an automated system for classifying curl patterns using computer vision and convolutional neural networks (CNNs). The goal is to build a model that recognizes hair patterns from an image, then uses this classification to recommend appropriate hair care products.

The system utilizes established systems for classifying curly hair:

- Andre Walker System: the primary system used to classify hair into four main categories each with subcategories.
- LOIS System and FIA System: alternative models that offer more comprehensive classification using additional features such as volume, density, shine, etc.

To collect various hair images, they used web scraping tools, as well as images taken from 30 volunteers. The images were then processed and converted to grayscale after that to binary images. By using “Active Shape Models” features were extracted and trained on a CNN model fed with 30x30 pixel images. The system was able to recognize curl patterns and recommend appropriate products based on the pattern.

Gaines, M. K., [10] reimagine curly hair classification using new and accurate geometric and mechanical measurements instead of traditional classifications based solely on external appearance and general appearance, such as the Andre Walker or L’Oréal systems.

The paper shows that traditional systems are insufficient to distinguish the differences between curly and coily hair types, which are a challenge users face when choosing appropriate products. Therefore, it aims to develop a new system based on precise geometric measurements such as the total length of the hair path (L), number of curls, pitch, diameter, and ellipticity (e), in addition to mechanical properties such as the tensile stress-strain relationship until the hair is stretched and breaks. It seeks to link these properties to hair type and appropriate products.

This study was conducted by collecting real hair samples from female volunteers at Spelman College with curly and coil hair. The researchers used precision instruments such as scanning electron microscopy (SEM), photogrammetry, tensile strength analyzer (TA), and dynamic mechanical analyzer (DMA) to measure hair shape, the number of bends per 3 cm (contours), the length-to-bend ratio (contour/pitch), and the stress-strain relationship. The results showed that curly and kinky hair have different properties that influence their performance and response to products. And they proposed a new quantitative classification system based on these geometric values.

Lasisi, T., Zaidi, [14] proposed that morphology is the science that studies the structure and shape of organisms, and the morphology of hair varies between individuals and people, and is often described in terms such as hair type, hair shape or hair texture (pattern). These descriptions may be inaccurate and depend on visual estimates instead of quantitative measurement, as this estimate may vary from person to person. A new measurement method has been proposed and developed to reduce human intervention. This method does not require high technical skills and is done through high-resolution imaging and automated image analysis. In this paper, the researchers developed a Python package called fibremorph, which extracts bending and eccentricity from longitudinal and transverse sections of hair images. This package does not need user programming and provides clear instructions for using it. Fibremorph is an essential tool that allows accurate measurement of hair specifications, where hair curvature is calculated scientifically instead of relying on visual observation.

The researchers collected a sample of 140 people of European and African origin, and their hair curves were quantifiably compared with the classification of hair patterns (straight, wavy, curly). Then the accuracy of this package was tested by simulating 100 samples, and the error rate was very low. The researchers adopted a protocol for hair stabilization and made cross-sections from it, where they used heat-resistant plastic sheets of polycaprolactone, heating plates, metal molds, cutting tools, and accurately represented the curvature of the hair through small samples (~3 mm) placed in a petri dish containing isopropanol alcohol (IPA)

Dealing with curly hair and its patterns requires accurate data on hair characteristics such as hair curvature and hair diameter. This paper provides a scientific basis for collecting and analyzing this data, which facilitates the training of AI models.

And the Reimagining Hair Science paper presents new standards for curly hair classification through geometric measurements. This paper supports this idea by providing a standardized measurement tool for these criteria.

[17] Makkar, J., Flores, presents a tool based on AI and computer vision to accurately measure and analyze hair characteristics to understand the changes that occur in hair throughout age, and the impact of hormones and genetic factors on hair follicles.

The researchers developed a computer vision methodology that recognizes overlapping hairs and measures them from images directly without the need to separate them manually, and performed a parallel treatment of several areas of a single image using a Sliding Window. In this study, samples taken from mice of different ages (from 21 days to 1.5 years) were collected, and the hair was cut from specific areas in the body of the mouse and saved for imaging. After that, the images were divided into 1024×1024 pixels to facilitate processing and reduce the computer load. The OBBDetection model was used to detect hairs and define

their boundaries, and the ResNet152 model was trained to recognize hair types with an accuracy of 96.9%.

As a result, they compared a hair classification (specific to humans) made manually and the classification using the tool in terms of length and percentage of each hair type and quantitative measurements such as area, circumference, color (RGB), and the percentage of dark color., the results showed that the classification and measurements resulting from the tool are highly compatible with the manual results. Finally, this study proves that the quantitative analysis of hair types gives strong indicators to understand hair development with age or health conditions, and explained how temporal changes in hair characteristics can be tracked under the influence of biological factors (maturity, hormones), which opens the way to study the effect of age or health status on the pattern of curly hair, orientation and color spectroscopy can be used to detect the chemical effect of products on color and shine in curly hair.

Erdogan, K., Acun, O., [11] introduces the KEBOT system, which utilizes artificial intelligence to accurately analyze hair before and after hair transplantation procedures, and provide surgeons with accurate reports to help in evaluating the procedure. The system consists of:

1. A 6-axis robot that moves around the patient's head.
2. A depth camera that generates a 3D model of the head.
3. A high-resolution RGB camera with a telecentric lens.
4. A dedicated light source to optimize the captured images.
5. A processing unit.
6. Artificial intelligence algorithms.

It aims to detect the number and location of follicles, calculate the number of hairs within each follicle, measure the diameter of each hair (in microns), calculate the hair coverage index, and generate visual and analytical reports for the physician.

The error rate of this system was measured to be less than 5% in most cases, and the results were very close to microscopic measurements.

What makes this study special is its usintelligence andhnologies such as YOLO, RetinaNet, and SegNet, its comprehensive coverage of all stages of transplantation (before and after), its combination of robotics and artificial intelligence, and providing accurate digital analysis to support the physician's decision.

However, its weakness is that it is designed for specific cases (men with AGA only) and does not cover different hair types or non-medical conditions. In addition, it is not intended for the end user, but rather for the physician only and there is no customization based on hair type or product.

Kim, Y., Kim, S., [16] this paper explains that there is no comprehensive system built with Artificial Intelligence (AI) that combines the diagnosis of scalp diseases and baldness due to data imbalance, in addition to the fact that the cost of developing Labels is very high because it needs specialists. Therefore, they presented a solution for a comprehensive diagnosis system based on AI that depends on the division of hair without stickers (Label-Free Segmentation), generating artificial images of rare cases (Diffuse IT-M), also extracting the features of hair (number, thickness) to assess the severity of baldness. In this paper, the researchers used a specialized database from AL-Hub containing four levels of severity (Good, Mild, Moderate, Severe), 6 scalp diseases such as Dandruff, Sebum, Erythema, Hair loss, baldness (Alopecia), and about 24,000 images for baldness cases only. They suggested a diagnostic system (Scalp Vision) based on two main indicators:

1. Hair Thickness – Hair Thickness.
2. Hair Count – Number of hairs.

To obtain the number and thickness of hairs as indicators of baldness.

And three basic steps:

1. Hair Segmentation Module, where the hair is divided from microscopic images.
2. Data Augmentation by converting images to a specific pathological condition via Image-to-Image Translation while maintaining hair details, in addition to generating images to expand the training data and improve classification accuracy
3. Classification & Severity Prediction, where scalp diseases are classified and the severity of baldness is estimated.

Therefore, the idea of Hair Thickness and Hair Count can be applied to curly hair to determine the pattern or thickness of the hair, and Image-to-Image Translation may be used to simulate different cases of hair conditions, such as wet, dry, conditioned with certain products and their response to them, and thence train an intelligent recommendation system

Meishvili, G., Clemoes, [15] presents a new hairstyle classification (Hairstyle Taxonomy) developed in collaboration with hairdressers, anthropologists, and technicians using frequent reviews with 30 evaluators from different backgrounds and have been repeatedly improved based on experiences and studies. The presented classification system is based on computer vision and deep learning that predicts the person's hairstyle from a single image using clear characteristics across different hair types, the current systems either classify hair directly or by parametric modelling, these methods may be accurate for straight or wavy hair, but they are still weak in representing complex hair patterns such as curly or gathered hairstyles such as cake or horsetail styles. This classification system contains 18 basic properties that describe the shape of the hair as a whole and divides the scalp into 8 regions, each region is characterized by distinctive characteristics such as tuft length, density, style type, and each hairstyle has 74 labels based on these characteristics. The study was based on two types of data, industrial data generated from the Wood et al library. 2021 In addition to the FairFace

Dataset, with images cut to be clear to hair, these images have been carefully selected to include different hair lengths and types, lighting angles and various positions, and proposed a visual classification (Graphical overview) of hairstyles divided into global attributes such as baby hair, hairline visibility, etc and regional attributes such as thickness, hair type, etc.

This paper reinforces the idea that the exact classification of curly hair patterns is essential before any recommendation of products, with a challenge that must be taken into consideration while analyzing images, which is the difficulty of recognizing hair in poor lighting or complex backgrounds.

Rosu, R. A., 1★, Saito, S., [12] presents a new method for representing and generating 3D hair strands from multi-view images in a geometrically accurate, realistic, and easily modifiable way. Previous papers have focused on recognizing hair type from static images, while the “Neural Strands model” emerges as an advanced analytical tool that enables accurate reconstruction of the full hair shape from multi-view images. This opens the door to artificial intelligence applications. The Neural Strands model not only recognizes individual hair, but also restores the full shape of the strands, even from different angles, and this is important for hair care applications because it analyzes the shape of each strand, its curvature, density, and growth direction. The model develops a system for generating realistic 3D hair strands from multi-angle images, using texture maps representing the shape and appearance of each strand. They use UNET (a U-shaped encoder-decoder network architecture) to generate a realistic final image (RGB + alpha). It was trained on real images, and two types of losses are used:

1. A geometric loss to approximate the hair shape.
2. A visual loss (rendering loss) to approximate the final appearance with real images.

This paper is considered as the first learning framework that combines hair shape and realistic appearance from multi-angle images, and it is usable in real time for generating realistic images of human hair.

Rosu, R. A., Wu, K., [13] rebuild 3D hair geometry from a single image using diffusion models which is a serious technical challenge especially for curly or complex hair, such as Afro hairstyles, due to the wide variety of shapes and the lack of realistic data.

The researchers proposed a new framework called DiffLocks to solve two main problems:

1. Lack of realistic 3D data.
2. Reconstruction of the fine details of each hair.

They created a massive database containing 40,000 (40 K) realistic and diverse 3D hair styles, which now serves as a reference for other researchers. The researchers used Blender to create 3D hair strands, and they trained their model using DINOv2 features to guide the generation process, giving the model the ability to understand the fine spatial details of each hair strand. The model requires no post-processing or additional random strand injections, but rather automatically generates hair containing fine details such as flyaway strands. It is

worth mentioning that the model was trained only on synthetic data, yet it successfully generalized to real images. The DiffLocks paper doesn't focus on product recommendations and damage analysis, but it does offer the ability to generate accurate curly hair images that can be used in model training.

Chapter 3 Proposed Approach (Methodology)

3.1 introduction

This chapter describes my research methodology and demonstrates steps followed to classify hair type and damage degree. The approach will use two different models, a DL model for hair classification and a ML used for analyzing SEM images for features extraction. My methodology pipeline will follow these steps:

- 1- data collection and preprocessing
- 2- feature selection step.
- 3- multiple DL and ML models will be trained and evaluated to achieve a better performance

3.2 Dataset Description

3.2.1 Hair Type Classification

A dataset found in Kaggle designed to classify various hair types. It includes high-quality images of individuals with diverse hair types. The dataset includes the major 4 types of hair and dreadlocks. The dreadlocks is not actually hair type but a hair style so it's have been excluded

Dataset Distribution

Hair Type	Number of Images
Straight	488
Wavy	330
Curly	514
Dreadlocks	443
Kinky	217
Total	1992

Table 1 – hair type dataset

3.2.2 SEM Images

The datadryad SEM dataset was used to evaluate the proposed methodology. Ana L'Oréal SEM dataset was selected to test the effectiveness of the proposed approach in analyzing hair damage.

Dataset Distribution

Dataset	Number of Images	Image Type
datadryad	47	Tif
L'Oréal	429	png

Table 2 – SEM datasets

Therefore, this research relay on two different datasets:

1. Plain images for hair type classification
2. SEM images for hair damage analysis.

3.3 Data Preprocessing

3.3.1 Hair Type Images Preparation

Data augmentation techniques were applied to increase image diversity and reduce overfitting such as:

1. A random resized crop with 0.80 - 1.0 scale and change width to height ratio from 0.75 -1.33 to prevent the model from saving the same hair position in all images.
2. A random horizontal flip by 50% probability
3. A random rotation
4. Change brightness 0.25, contrast 0.25, saturation 0.15, and hue 0.02)
5. Random erasing to prevent the model from relying on only one part of the image.

3.3.2 SEM Image Preparation

For SEM images, the processing was different because the goal wasn't just direct classification from the raw image but extracting specific properties. Therefore, the images were manipulated to facilitate the extraction of the desired properties and to analyze the hair surface structure more clearly.

3.4 Deep Learning Model

To classify hair, type a CNN used to analyze images and extract complex features, an EfficientNet-B0 model was chosen. The reason for choosing EfficientNet-B0 is its suitability for image processing, and it achieves a good balance between performance and accuracy, it's also known for its efficiency in extracting important image features without much complexity. The model can capture the visual differences between various hair types, such as strand shape, curl level, and overall hair density.

3.4.1 Training Process

After preparing images, dataset was splitted for:

1. training set
2. validation set
3. testing set

3.4.2 Classification Task

The goal is to classify images to one of these classes

1. Straight
2. Wavy
3. Curly
4. Kinky

3.5 Machine Learning Model for SEM Analysis

This part of research focuses on analyzing SEM images. SEM images provide a detailed view of the hair surface, which helps in studying hair condition and identifying characteristics that may be related to damage. SEM provides details that are not visible in regular images, so it helps in examining the hair's structure more deeply.

3.5.2 Feature Extraction

The steps were implemented based on the referenced research paper [4]

Step 1: Convert image to grayscale

Step 2: Remove SEM label by scanning image and crop white label area if found

Step 3: Detect hair orientation, apply Gaussian blur and Otsu threshold then detect largest component and estimate angle using PCA

Step 4: Rotate image horizontally

Step 5: Define ROI to extract central region

Step 6: Normalize brightness

Step 7: Generate histogram

Step 8: Find peak brightness

Step 9: Calculate $0.02H$

Step 10: Compute lead and lag

Step 11: Calculate Tailing Factor

Since expert annotation was not available, an unsupervised learning approach (K-means clustering) to group the data into relevant clusters.

3.6 Proposed Framework

Proposed frameworks consist of two paths.

First Path: Input plain hair image and predict hair type.



Fig.8. Hair type classification

Second path: Input SEM image, extract tailing factor and predict damage



Fig. 9. SEM analysis

3.7 Evaluation Metrics

After training the models, performance was evaluated using a set of appropriate metrics like accuracy, precision and F1-score. These metrics help to understand how successful the model is at classification to specify whether the model is working well or not.

3.8 Summary

This chapter presents the proposed methodology used in the research to analyze hair through two different approaches. The first approach relied on a DL model to classify hair types using plain images, while the second approach relied on ML to analyze SEM images after extracting the tailing factor feature. The results obtained from applying this methodology will be presented in the following chapter.

Chapter 4 :Experimental Results and Discussion

4.1 Introduction

In this chapter I will present results after applying the proposed methodology which was explained in the last chapter and a number of experiments were implemented to achieve best accuracy

4.2 Curly Hair Type Classification

A number of experiments were implemented to improve the performance of the proposed model and achieve best accuracy. First, the model was trained using default settings and then, parameters such as learning rate and batch size were modified, and data augmentation techniques were employed to enhance the model's performance.

4.2.1 Experiment 1

The learning rate was set to **0.001**, number of epochs **8** and the batch size was **16**. The model achieved an accuracy of **87%**

Best Val Acc: 0.8712

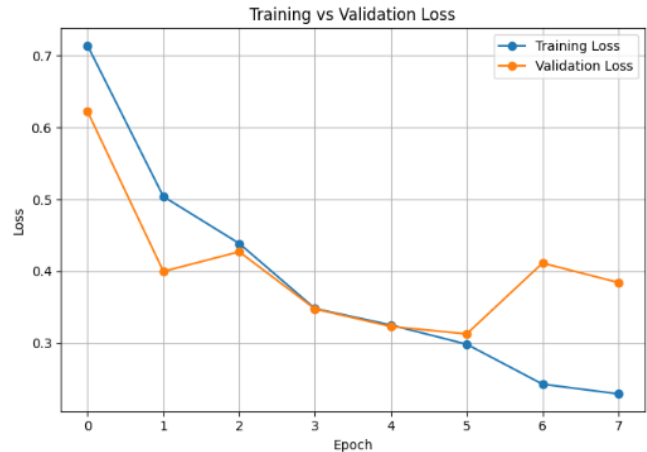
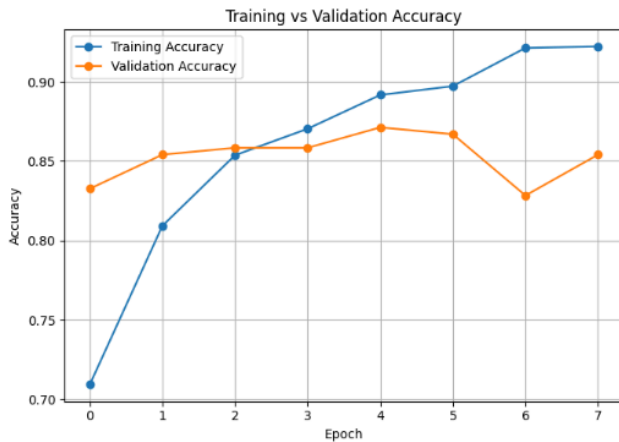


Fig. 10. Hair Type Classification - Experiment 1

Model: efficientnet_b0

Accuracy: 0.8670

Precision: 0.8748

Recall: 0.8670

F1-score: 0.8680

Classification Report:

	precision	recall	f1-score	support
curly	0.91	0.81	0.86	77
kinky	0.77	0.91	0.83	33
straight	0.94	0.89	0.92	73
wavy	0.79	0.90	0.84	50
accuracy		0.87		233
macro avg	0.85	0.88	0.86	233
weighted avg	0.87	0.87	0.87	233

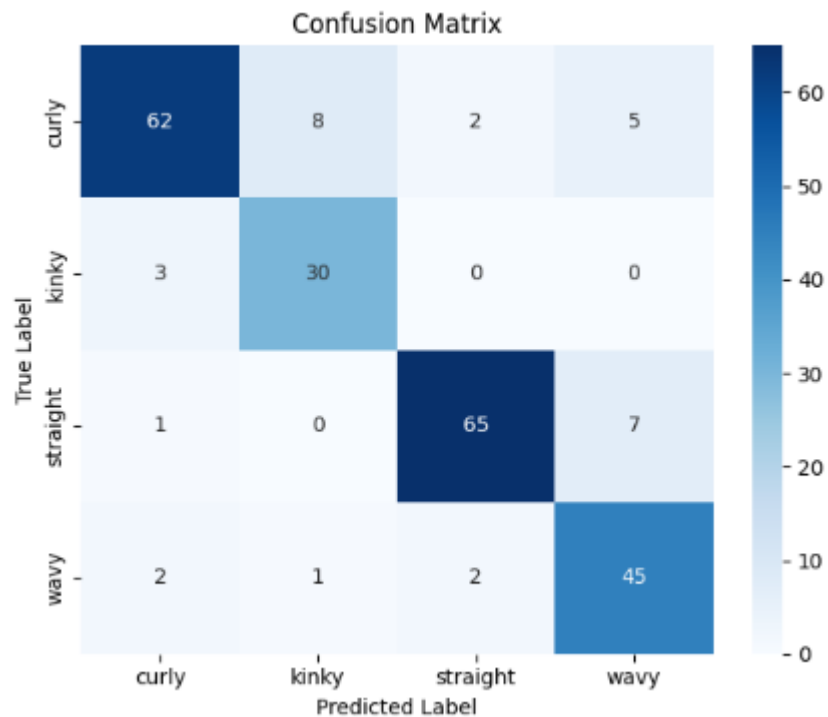


Fig. 11. Hair Type Classification - Experiment 1 – confusion matrix

4.2.2 Experiment 2

The learning rate was set to **0.0001**, number of epochs **8** and the batch size was **16**. The model achieved an accuracy of **91%**.

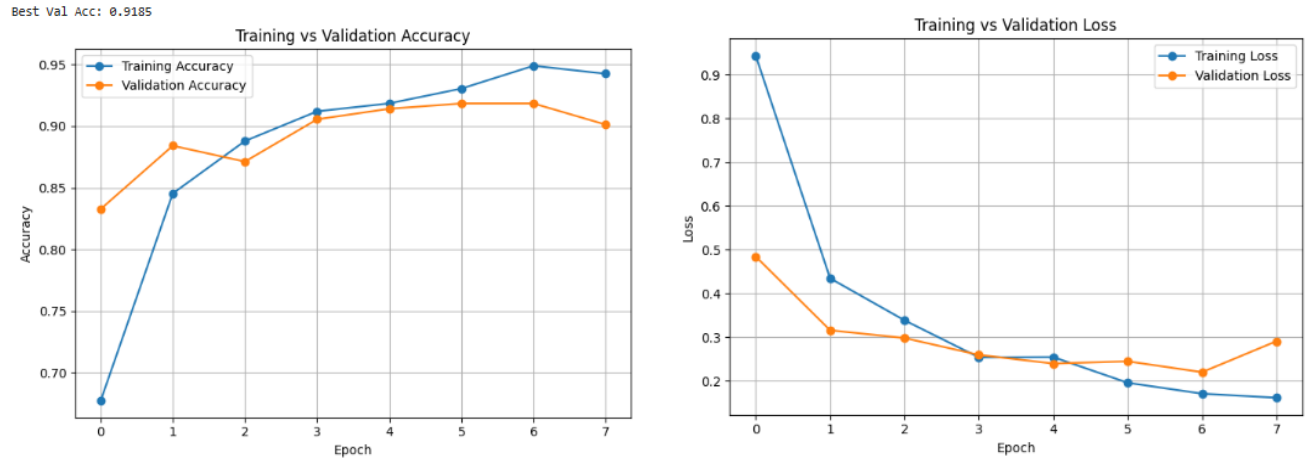


Fig. 12. Hair Type Classification - Experiment 2

Model: efficientnet_b0

Accuracy: 0.9142

Precision: 0.9186

Recall: 0.9142

F1-score: 0.9141

Classification Report:

	precision	recall	f1-score	support
curly	0.97	0.87	0.92	77
kinky	0.82	0.97	0.89	33
straight	0.91	0.97	0.94	73
wavy	0.91	0.86	0.89	50
accuracy		0.91		233
macro avg	0.90	0.92	0.91	233
weighted avg	0.92	0.91	0.91	233

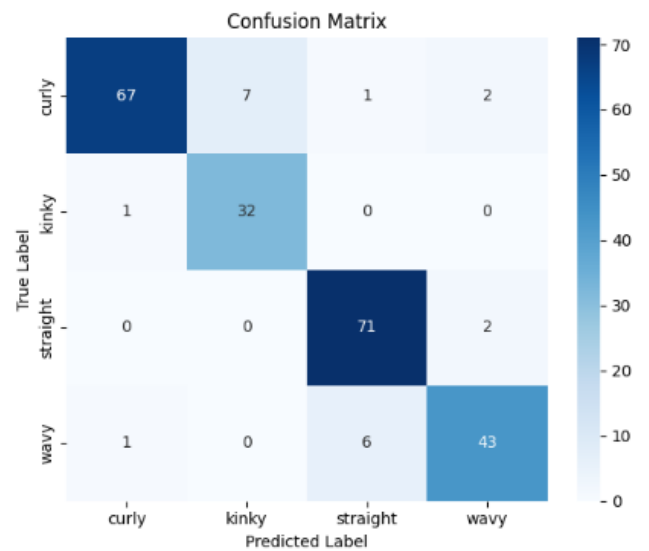
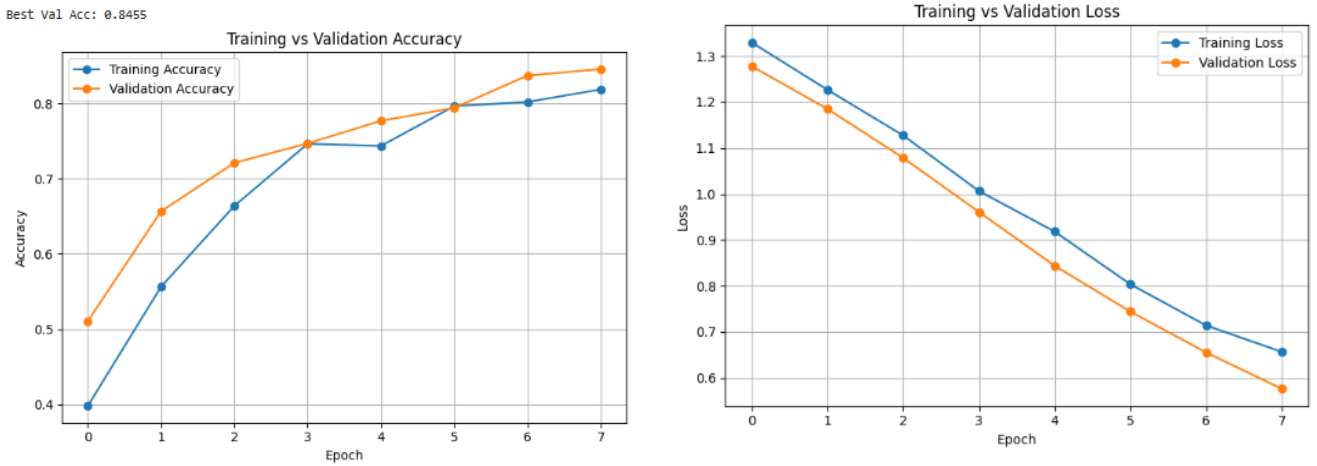


Fig. 13. Hair Type Classification - Experiment 2 – confusion matrix

4.2.3 Experiment 3

The learning rate was set to **0.00001**, number of epochs **8** and the batch size was **16**. The model achieved an accuracy of **84%**

Fig. 14. Hair Type Classification - Experiment 3



Model: efficientnet_b0

Accuracy: 0.8412

Precision: 0.8453

Recall: 0.8412

F1-score: 0.8364

Classification Report:

	precision	recall	f1-score	support
curly	0.80	0.92	0.86	77
kinky	0.89	0.73	0.80	33
straight	0.86	0.95	0.90	73
wavy	0.86	0.64	0.74	50
accuracy	0.84			233
macro avg	0.85	0.81	0.82	233
weighted avg	0.85	0.84	0.84	233

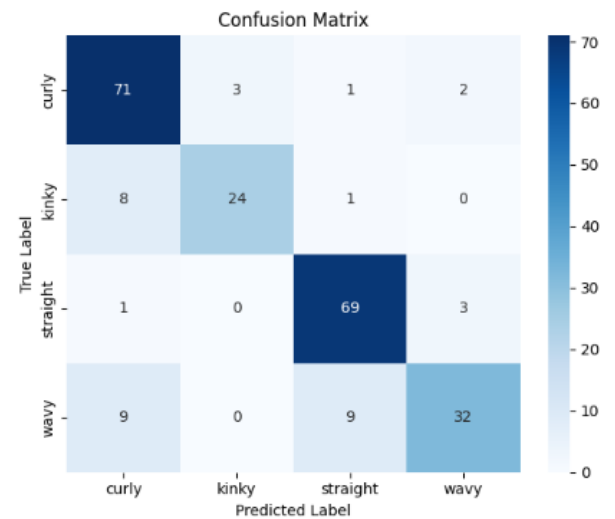


Fig. 15. Hair Type Classification - Experiment 3 – confusion matrix

4.2.4 Experiment 4

The learning rate was set to **0.0001**, number of epochs **15** and the batch size was **16**. An Early Stopping was added. The model achieved an accuracy of **93%**

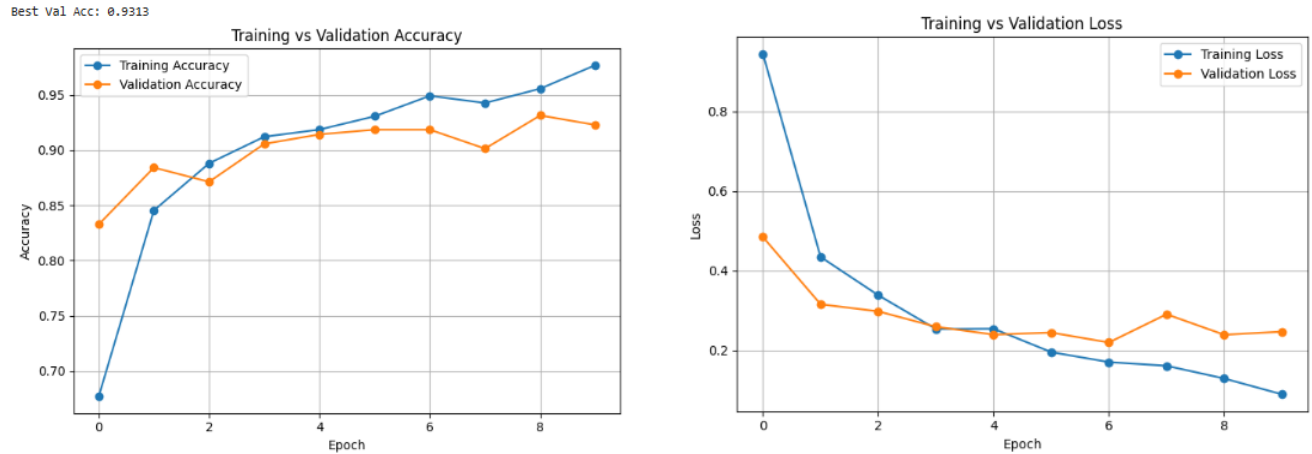


Fig. 16. Hair Type Classification - Experiment 4

Model: efficientnet_b0

Accuracy: 0.8841

Precision: 0.8876

Recall: 0.8841

F1-score: 0.8833

Classification Report:

	precision	recall	f1-score	support
curly	0.91	0.87	0.89	77
kinky	0.78	0.97	0.86	33
straight	0.92	0.95	0.93	73
wavy	0.88	0.76	0.82	50
accuracy		0.88		233
macro avg	0.87	0.89	0.88	233
weighted avg	0.89	0.88	0.88	233

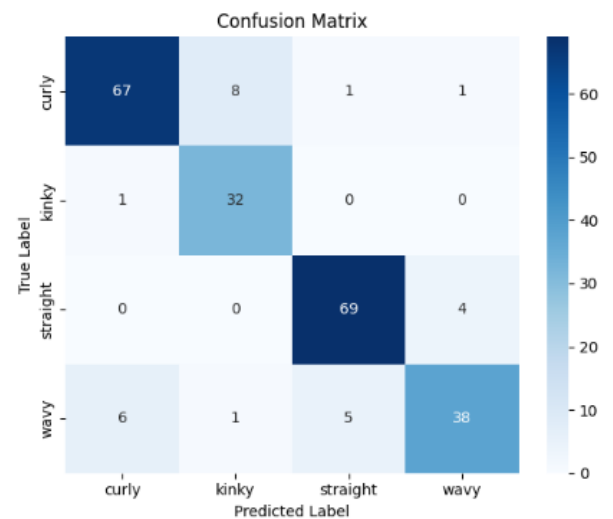


Fig. 17. Hair Type Classification - Experiment 4 – confusion matrix

4.2.5 Experiment Summary

Experiment	Learning Rate	Batch Size	number of epochs	Accuracy
Exp 1	0.001	16	8	87%
Exp 2	0.0001	16	8	91%
Exp 3	0.00001	16	8	84%
Exp 4	0.0001	16	15	93%

Table 3 – Hair Type Classification - Experiment Summary

The results showed that tuning these parameters significantly improved the model's performance, achieving optimal accuracy in the final experiment. These experiments proved the accuracy of the proposed model and its ability to achieve good results supporting its use in hair type classification.

4.3 SEM Analysis

SEM images provide a detailed view of the hair surface, which helps in studying hair condition and identifying characteristics that may be related to damage. It provides details that are not visible in regular images, so it helps in examining the hair's structure more deeply. The tailing factor is used as the main feature to represent the level of hair surface damage, the higher its value, the greater the damage. To validate the effectiveness of the proposed methodology, a number of experiments were implemented on two different SEM datasets. The first Dryad dataset, which also used in Fanny Chu [4] research to ensure consistency.

4.3.1 Dryad Dataset Experiments

Experiment 1 - RIO 75% - FILTER_MAX_4000X FALSE

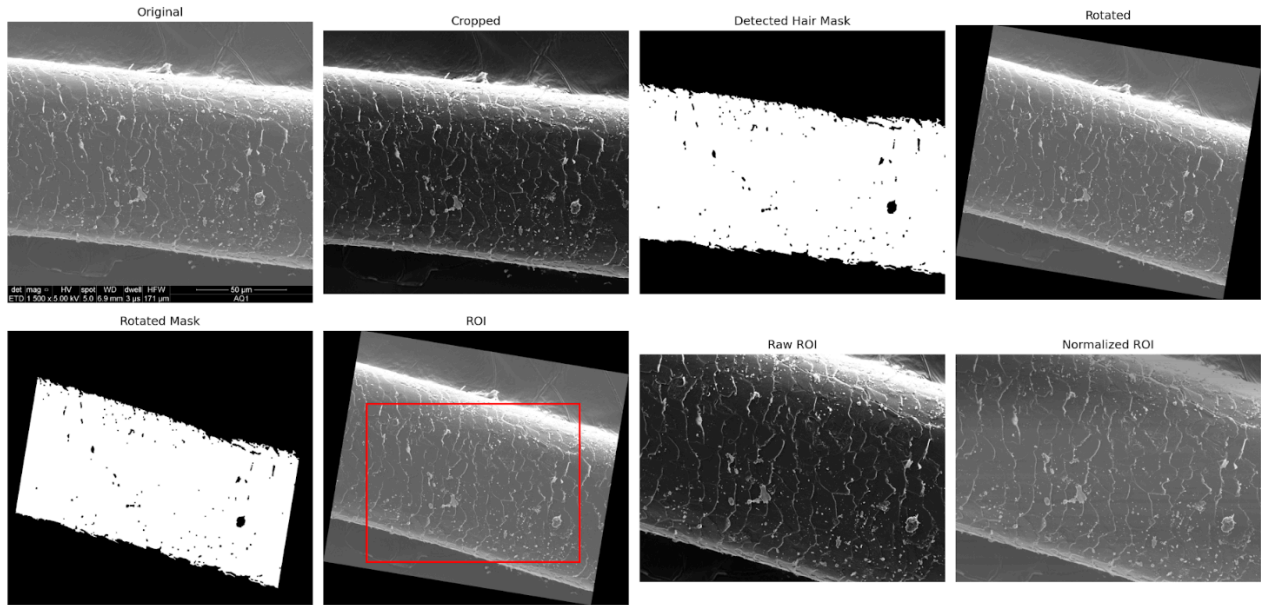


Fig. 18. SEM Analysis - Experiment 1- RIO

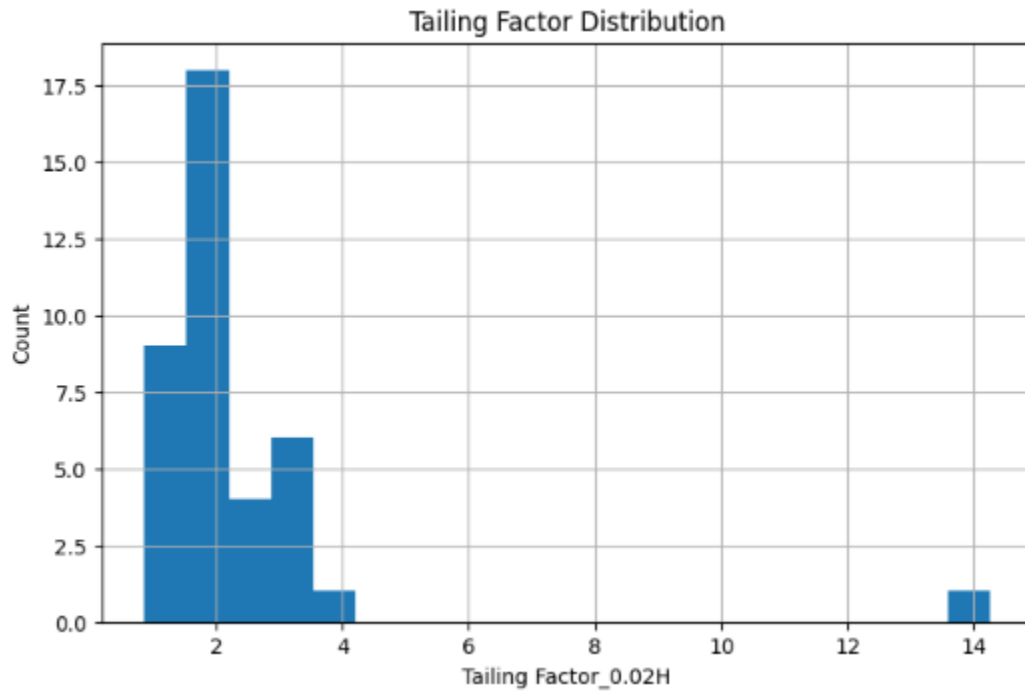


Fig. 19. SEM Analysis - Experiment 1 – Tailing factor distribution

Experiment 2 - RIO 70% - FILTER_MAX_4000X FALSE

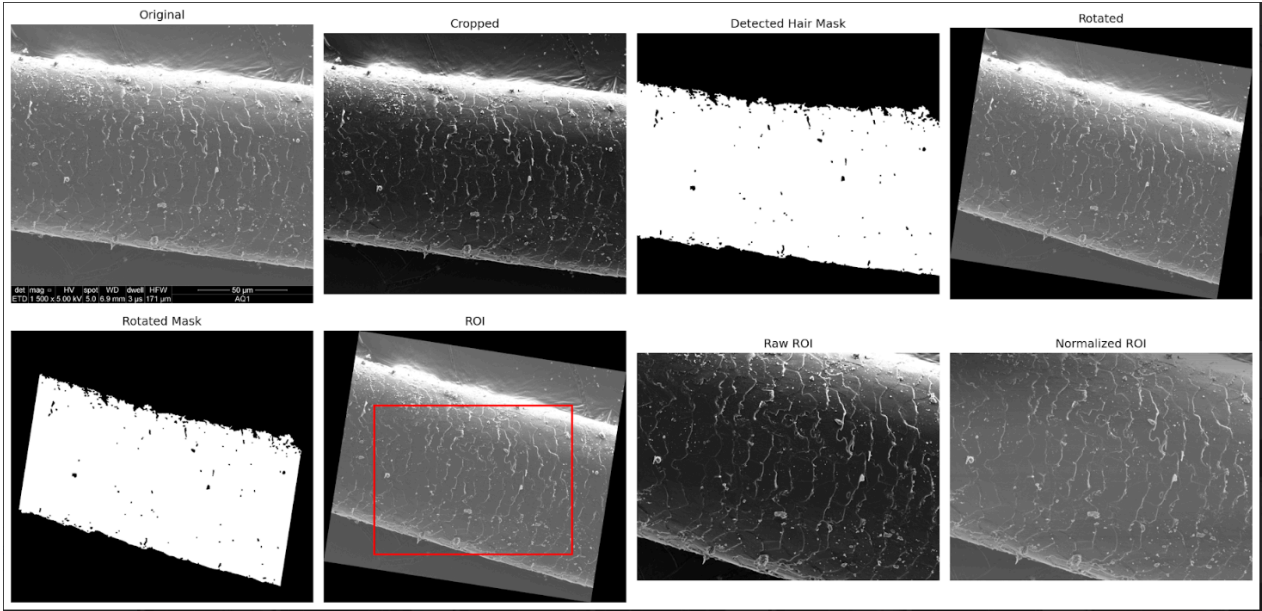


Fig. 20. SEM Analysis - Experiment 2- RIO

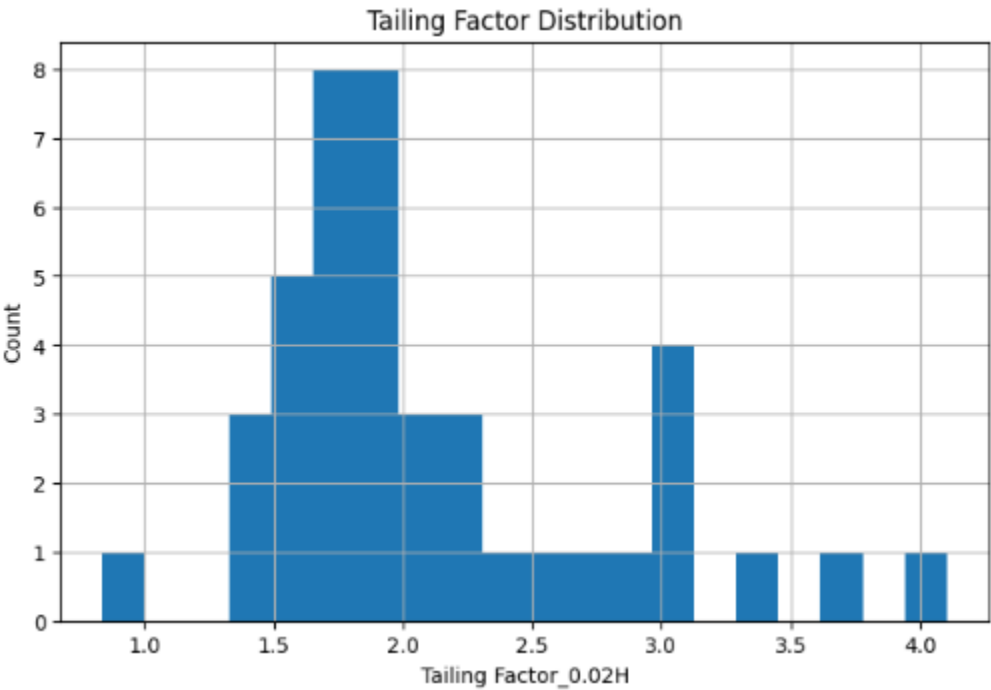


Fig. 21. SEM Analysis - Experiment 2 – Tailing factor distribution

Experiment 3 - RIO 50% - FILTER_MAX_4000X FALSE

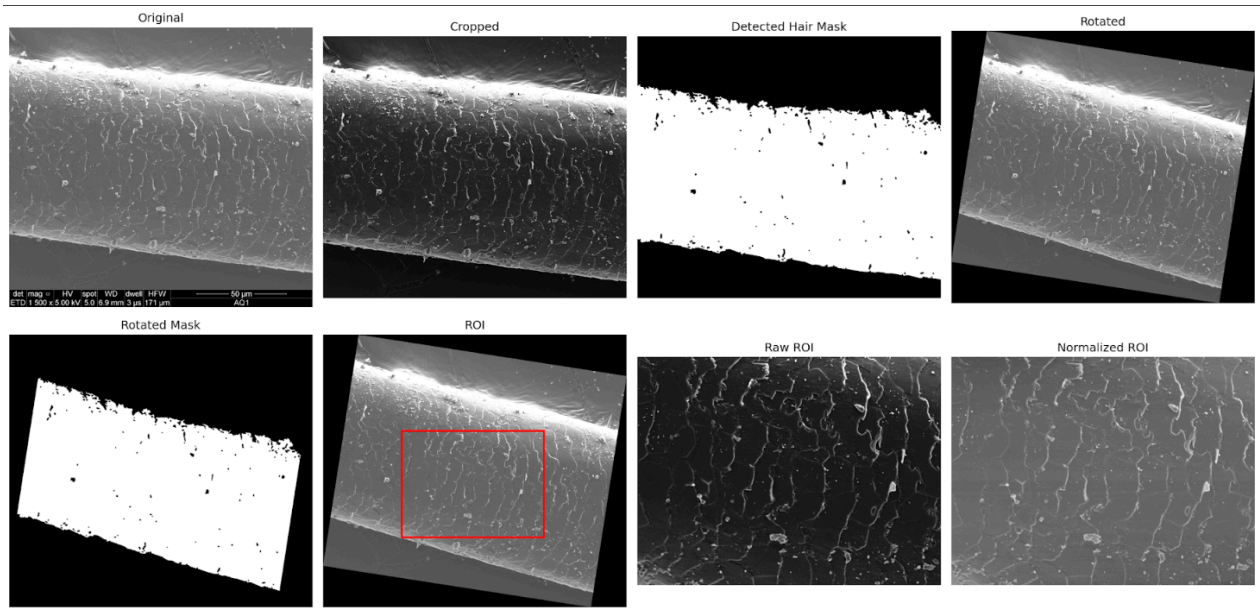


Fig. 22. SEM Analysis - Experiment 3- RIO

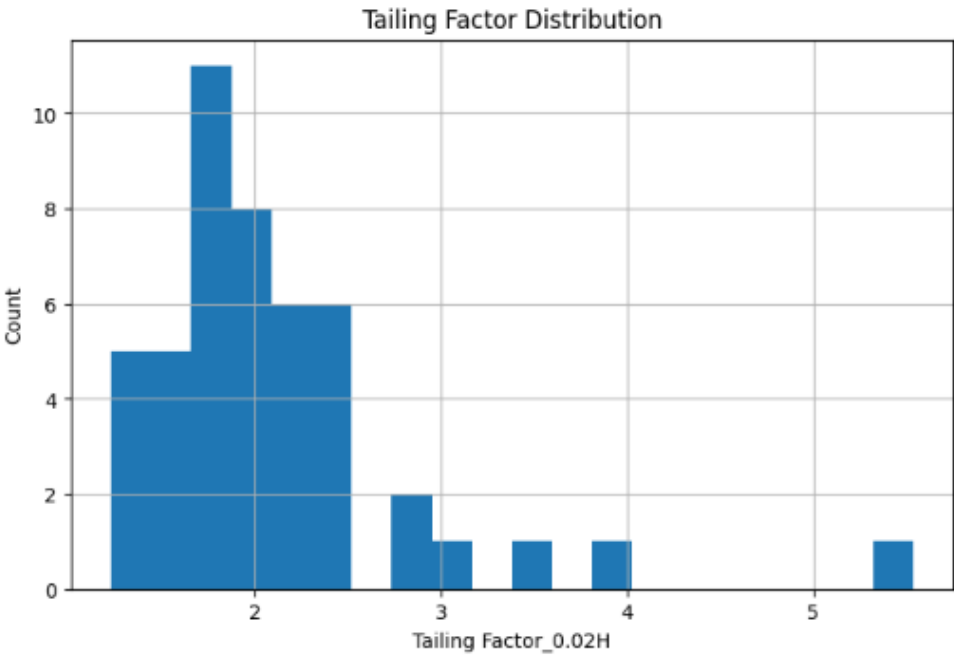


Fig. 23. SEM Analysis - Experiment 3 – Tailing factor distribution

Experiment 4 - RIO 40% - FILTER_MAX_4000X FALSE

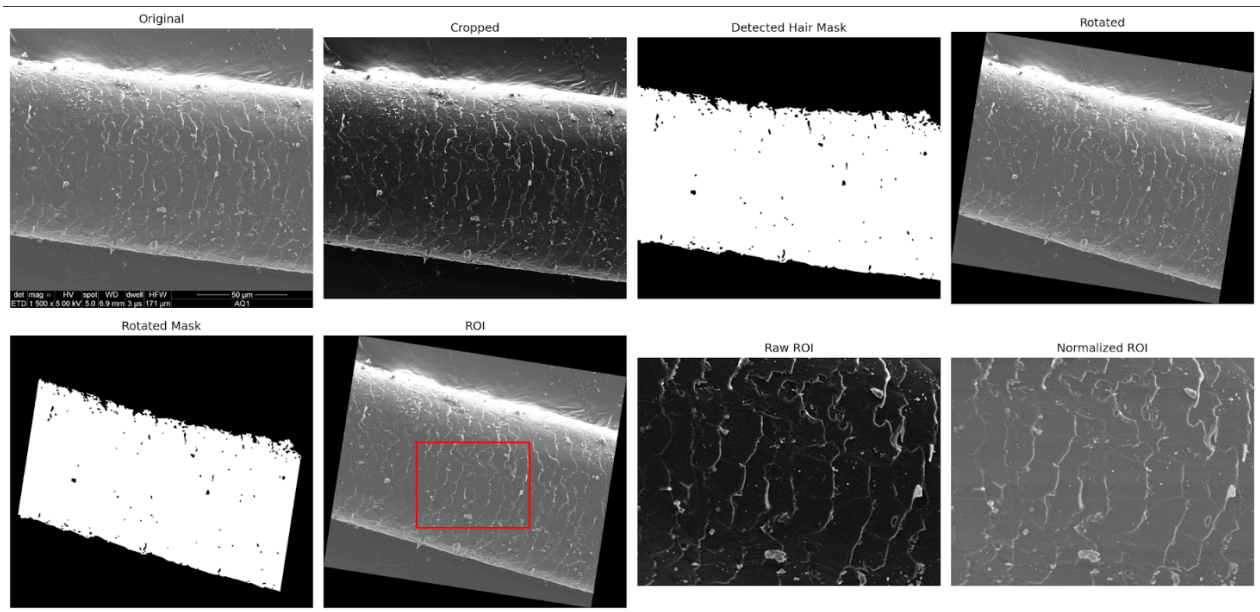


Fig. 24. SEM Analysis - Experiment 4- RIO

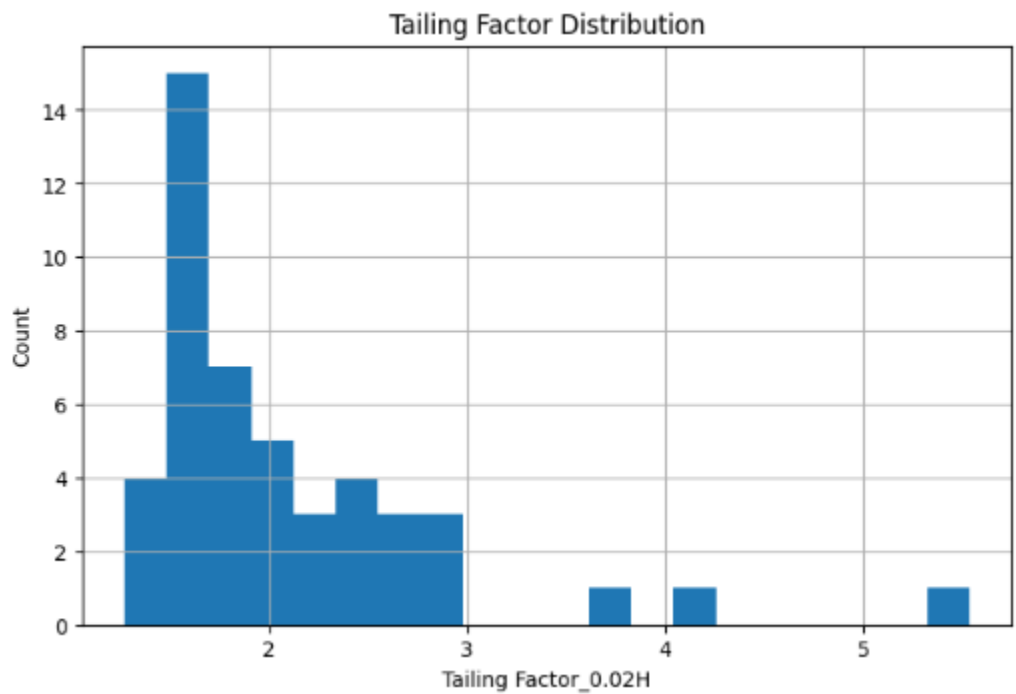


Fig. 25. SEM Analysis - Experiment 4 – Tailing factor distribution

Experiment 5 - RIO 40% FILTER_MAX_4000X TRUE

As paper proposed only for images < 4000X

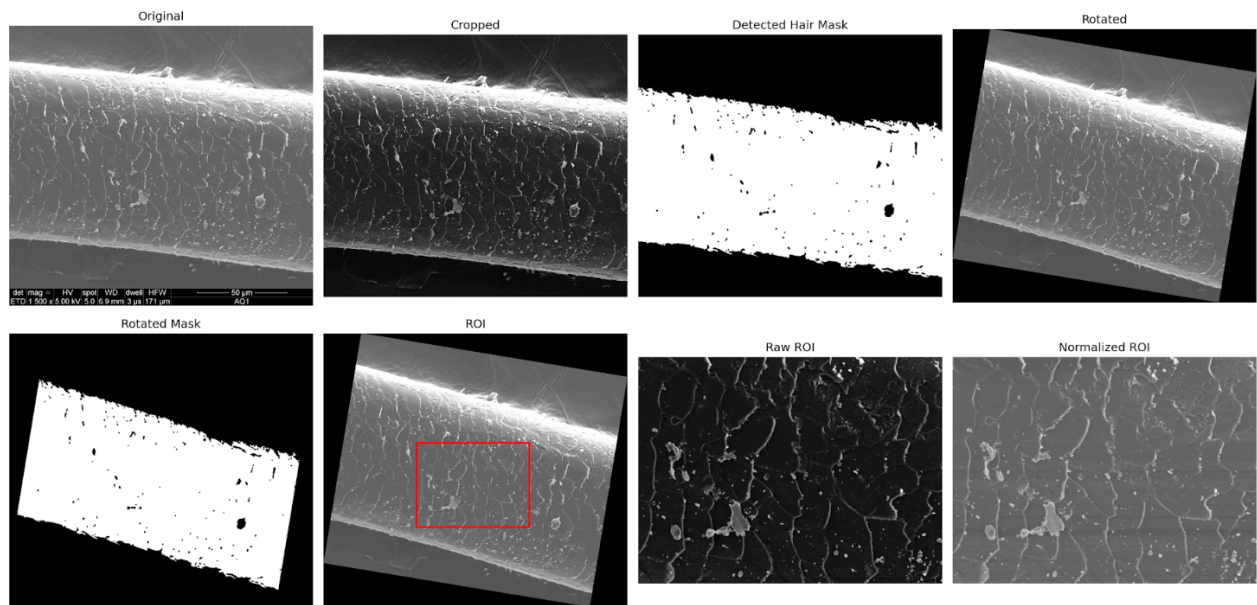


Fig. 26. SEM Analysis - Experiment 5- RIO

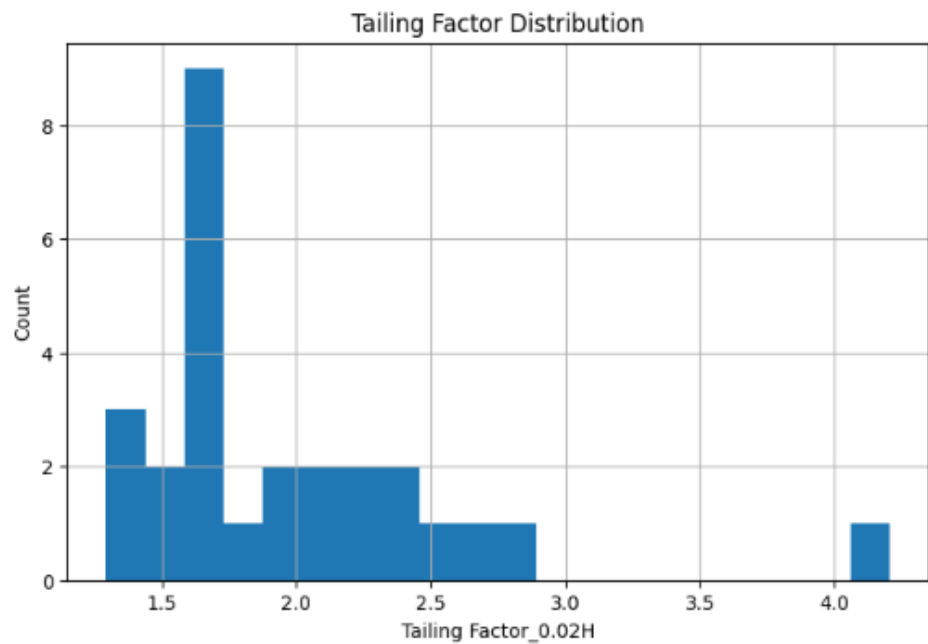


Fig. 27. SEM Analysis - Experiment 5 – Tailing factor distribution

4.3.2 L'Oréal Dataset Experiment

In addition to Dryad dataset, the methodology was evaluated using the L'Oréal dataset. This experiment was implemented to verify the effectiveness of the proposed method on a different dataset. The results demonstrated that the proposed approach performed effectively on both datasets.

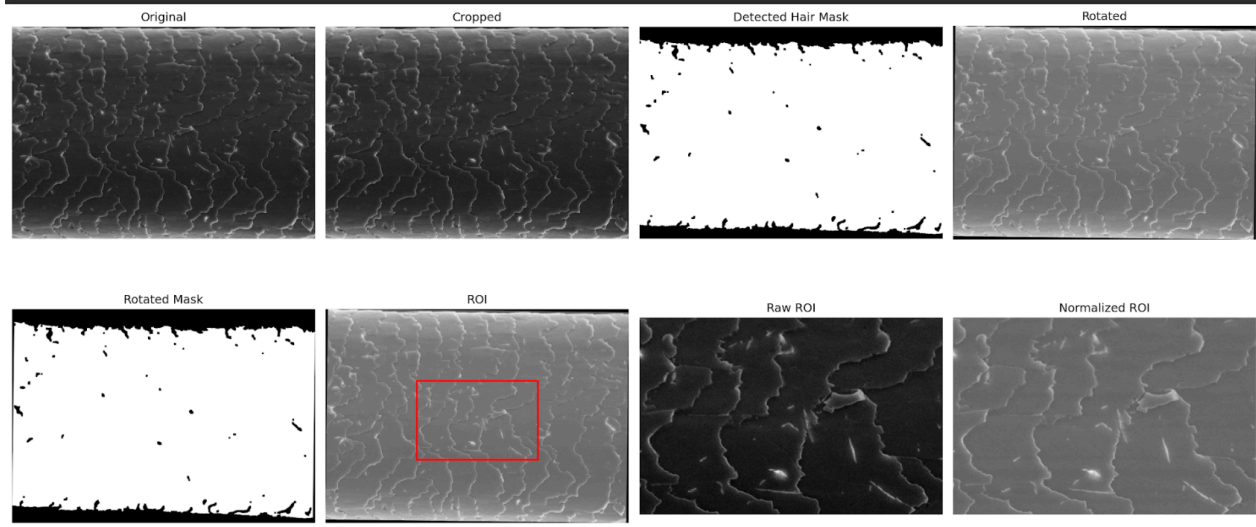


Fig. 28. SEM Analysis – L'Oréal RIO

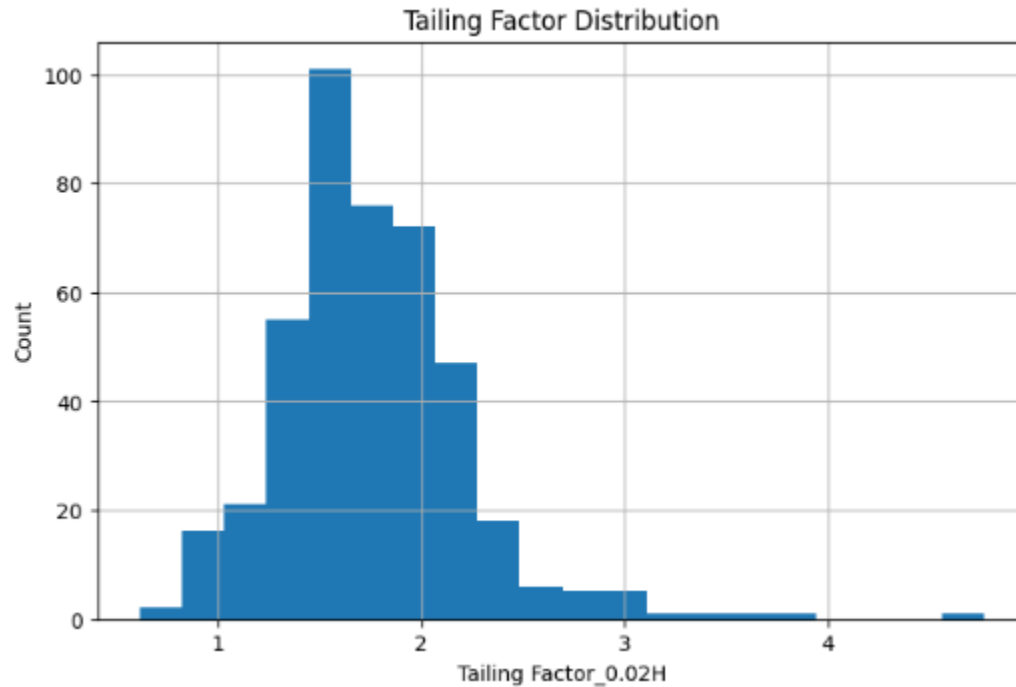


Fig. 29. SEM Analysis – L'Oréal Tailing factor

Since expert annotation was not available, an unsupervised learning approach (K-means clustering) to group the data into relevant clusters.

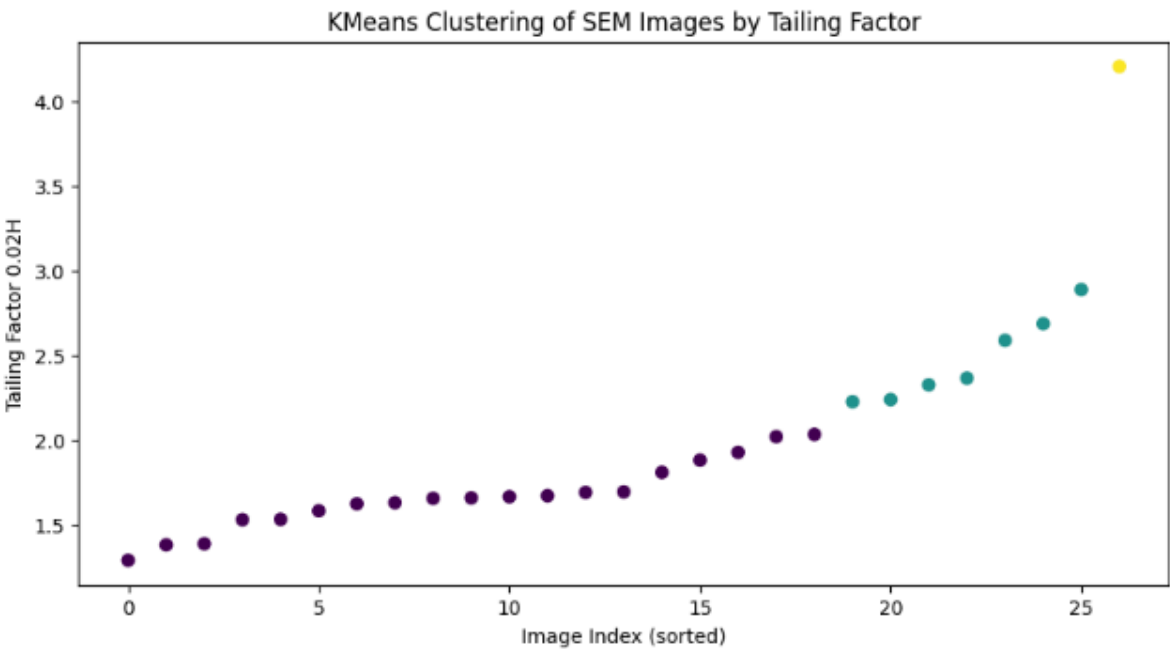


Fig. 30. SEM Analysis – Kmeans clustering by Tailing factor

Cluster Summary:

	count	min	max	mean
cluster_ordered				
0	19	1.291667	2.034483	1.667799
1	7	2.227273	2.888889	2.475220
2	1	4.205882	4.205882	4.205882

Chapter 5 Conclusion and Future Work

5.1 Conclusion

In this research, an AI-based framework for analyzing and classifying curly hair using two different methods was proposed. The first method relied on DL with the EfficientNet-B0 model to classify hair types from plain images, while the second method used SEM image analysis with ML techniques to analyze hair damage.

Number of experiments were conducted to improve the performance of the models used. The results showed that the EfficientNet-B0 model performed well in classifying hair types, and the SEM image analysis results show the proposed method's ability to distinguish between different hair damage levels.

Therefore, the results proved that the proposed framework is effective in analyzing and classifying hair and can contribute to the development of intelligent systems that provide appropriate hair care recommendations.

5.2 Future Work

This work can be improved; we can collaborate with a hair expert to validate the model results and refine the proposed recommendation system. The hair type derived from the DL model, along with the damage percentage extracted from SEM image analysis, can be used to provide accurate and personalized recommendations for each individual case.

References

- [1] AI for Non Techies. (2024, February 29). *Differences between AI, ML, DL, GenAI & LLMs*. Medium.
<https://ai-for-non-techies.medium.com/differences-between-ai-ml-dl-genai-llms-188340a35407>
- [2] IBM. What are convolutional neural networks (CNNs)? IBM Think. Available from:
<https://www.ibm.com/think/topics/convolutional-neural-networks>
- [3] Elsevier. (n.d.). *Artificial Intelligence*. In *Encyclopedia of Information Systems*. ScienceDirect.
<https://www.sciencedirect.com/science/chapter/referencework/abs/pii/B012227055X007793>
- [4] [26] Chu F, Anex DS, Jones AD, Hart BR. Automated analysis of scanning electron microscopic images for assessment of hair surface damage. *Royal Society Open Science*. 2020;7:191438. Available from: <https://doi.org/10.1098/rsos.191438>
- [5] IBM. What is k-means clustering? IBM Think. Available from:
<https://www.ibm.com/think/topics/k-means-clustering>
- [6] Muhammad UR, Svanera M, Leonardi R, Benini S.
Hair detection, segmentation and hairstyle classification in the wild. *Image and Vision Computing*. 2018.
- [7] Dertli B, Koklu M. Classification of hair types with deep learning methods. 5th International Conference on Trends in Advanced Research (ICTAR). Konya, Turkey; 2025.
- [8] Daniels, G., Tamburic, S., Benini, S., Randall, J., Sanderson, T., & Savardi, M. (2021). Artificial Intelligence in hair research: A proof-of-concept study on evaluating hair assembly features. *International Journal of Cosmetic Science*, 43(4), 405–418.
<https://doi.org/10.1111/ics.12706>
- [9] Arabnia, H. R., Deligiannidis, L., & Tinetti, F. G. (2019). *Image processing, computer vision, and pattern recognition*. 2018 Worldcomp International C.
- [10] Gaines, M. K., Page, I. Y., Miller, N. A., Greenvall, B. R., Medina, J. J., Irschick, D. J., ... & Crosby, A. J. (2023). Reimagining hair science: a new approach to classify curly hair phenotypes via new quantitative geometric and structural mechanical parameters. *Accounts of chemical research*, 56(11), 1330-1339.

- [11] Erdogan, K., Acun, O., Kucukmanisa, A., Duvar, R., Bayramoglu, A., & Urhan, O. (2020). KEBOT: an artificial intelligence based comprehensive analysis system for FUE based hair transplantation. *IEEE Access*, 8, 200461–200476. <https://doi.org/10.1109/access.2020.3035637>
- [12] Rosu, R. A., 1★, Saito, S., Wang, Z., Wu, C., Behnke, S., Nam, G., University of Bonn, Carnegie Mellon University, & Reality Labs Research. (2022). Neural Strands: Learning hair geometry and appearance from Multi-View images. *Neural Strands: Learning Hair Geometry and Appearance From Multi-View Images*. https://radualexandru.github.io/neural_strands
- [13] Rosu, R. A., Wu, K., Feng, Y., Zheng, Y., & Black, M. J. (2025). DiffLocks: Generating 3D Hair from a Single Image using Diffusion Models. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 10847-10857).
- [14] Lasisi, T., Zaidi, A. A., Webster, T. H., Stephens, N. B., Routch, K., Jablonski, N. G., & Shriver, M. D. (2021). High-throughput phenotyping methods for quantifying hair fiber morphology. *Scientific reports*, 11(1), 11535.
- [15] Meishvili, G., Clemoes, J., Hewitt, C., Hosenie, Z., Xiao, X., de La Gorce, M., ... & Wilczkowiak, M. (2024, December). Hairmony: Fairness-aware hairstyle classification. In *SIGGRAPH Asia 2024 Conference Papers* (pp. 1-11).
- [16] Kim, Y., Kim, S., Moon, H., Yu, Y., & Noh, J. (2024). Scalp Diagnostic System With Label-Free Segmentation and Training-Free Image Translation. *arXiv preprint arXiv:2406.17254*.
- [17] Makkar, J., Flores, J., Matich, M., Duong, T. T., Thompson, S. M., Du, Y., ... & Driskell, R. R. (2025). Deep Hair Phenomics: Implications in Endocrinology, Development, and Aging. *Journal of Investigative Dermatology*, 145(4), 800-811.
- [18] Massey, L. (2011). *Curly Girl: The Handbook* (2nd ed.). Workman Publishing Company.
- [19] Robbins, C. R. (2012). *Chemical and physical behavior of human hair* (5th ed.). Springer. <https://doi.org/10.1007/978-3-642-25611-0>
- [20] Cline, A. (2020). Moisture vs. protein balance: Why your curly hair needs both. *NaturallyCurly*. Retrieved from <https://www.naturallycurly.com>
- [21] Hill K, Adesola A, Dangal P, Washington G, Burge L. Towards creation of a curl pattern recognition system.

International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV); 2018.

Datasets

- [1] Hair Type Dataset: <https://www.kaggle.com/datasets/kavyasreeb/hair-type-dataset>
- [2] Automated analysis of scanning electron microscopic images for assessment of hair surface damage: <https://datadryad.org/dataset/doi:10.5061/dryad.ttdz08kt4>
- [3] L'Oréal_SEM_Hair_2024: https://people.cmm.minesparis.psl.eu/files/databases/Loreal_SEM_Hair_2024/