

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/366613168>

# Survey-based Machine learning approaches to diagnosis of hair fall disorder in Bangladeshi Community

Conference Paper · October 2022

DOI: 10.1109/ICCCNT54827.2022.9984226

---

CITATION

1

---

READS

761

5 authors, including:



**Mst. Farhana Khatun**

Daffodil International University

4 PUBLICATIONS 6 CITATIONS

SEE PROFILE



**Moshfiqur Rahman Ajmain**

Daffodil International University

8 PUBLICATIONS 9 CITATIONS

SEE PROFILE



**Sharun Akter Khushbu**

Daffodil International University

52 PUBLICATIONS 261 CITATIONS

SEE PROFILE



**Nushrat Jahan Ria**

Daffodil International University

19 PUBLICATIONS 49 CITATIONS

SEE PROFILE

# Survey-based Machine learning approaches to diagnosis of hair fall disorder in Bangladeshi Community

Mst. Farhana Khatun  
*Dept. of CSE*  
 Daffodil International University  
 Dhaka, Bangladesh  
 farhana15-14304@diu.edu.bd

Moshfiqur Rahman Ajmain  
*Dept. of CSE*  
 Daffodil International University  
 Dhaka, Bangladesh  
 moshfiqur15-14090@diu.edu.bd

Sharun Akter Khushbu  
*Dept. of CSE*  
 Daffodil International University  
 Dhaka, Bangladesh  
 sharun.cse@diu.edu.bd

Nushrat Jahan Ria  
*Dept. of CSE*  
 Daffodil International University  
 Dhaka, Bangladesh  
 nushratia.cse@diu.edu.bd

Sheak Rashed Haider Noori  
*Dept. of CSE*  
 Daffodil International University  
 Dhaka, Bangladesh  
 drnoori@daffodilvarsity.edu.bd

**Abstract**—Hair symbolizes the beauty of women and men. All of us are jealous of our hair. We lose hair at a young age due to some mistakes or irregularities. Lots of men and women all over the world are suffering from hair falling and the number of females is suffering growing per year. Genetically, dandruff, allergy and stress are the major problems for falling hair. We are doing this research survey for helping people. This study is representing two things. First of all, we are findings how many reasons are involved in hair fall. Another thing is we train our dataset with machine learning algorithms to find out the accuracy. Machine learning technologies have rapidly evolved to analyze survey datasets. SVM, Logistic Regression, Naive Bayes, Decision Tree, Random Forest, K-nearest Neighbor and XGBoost algorithms for performance comparison. The experimental results indicated that XGBoost had the best performance, with an accuracy of 92.62%.

**Index Terms**—Hair fall, Hair damage, Hair Fall Survey Analysis, Machine learning Algorithm, XGBoost Classification

## I. INTRODUCTION

One of the regular concerns of people seen is excessive hair fall. Every day some hair is falling from our head. But if the hair fall is a little extra, it leads us to anxiety and depression. To get rid of excess hair fall we are approaching experts. One of the questions in our mind is why the extra hair is falling out. Like our lives, hair cells also have a life cycle of their own. Its growth phase, persistence and subsequent shedding. But sometimes that life cycle can be disrupted. Several problems such as stress, poor lifestyle, lack of balanced diet, not sleeping properly or insufficient sleep can also cause hair fall. Vitamin B12, vitamin D3 and iron deficiency or anemia, hormonal imbalance or insulin resistance, PCOS (polycystic ovary syndrome, thyroid deficiency or increased stress and menopause, illness or pregnancy) can cause hair fall. In order to remove the deficiency and imbalance of hair on the head,

some rules must be followed such as - Nuts, dates, figs, grain seeds, colorful vegetables and green-yellow-orange-red fruits must be kept in the diet. Foods like curd, eggs, meat, cheese and milk should be eaten to meet the protein requirement. Also, we can make a diet plan and a routine with the advice of a nutritionist. When we see a problem, we find a solution. Here I will discuss tips for men and women on how to reduce their hair fall & proper treatment at the home. What do we know? Losing hair reduces self-confidence. Both women and men suffer from hair fall. However, even if women have the opportunity to have their hair proper care, men do not have the same opportunity. As a result, hair fall due to lack of proper care. Go bald at a young age. Vitamin E - Increases blood circulation in the scalp, and Vitamin E ensures the healthy color of hair. Do you know how much hair we shed every day? The answer is one hundred. Usually, we don't think about it every day, but when we lose a little more hair, we get worried. We can take care of our hair at home. Some notable ones are – fenugreek, neem leaf juice, aloe vera gel, coconut milk, henna and mustard oil, onion juice, egg yolk, oil, olive oil, cumin and honey. According to one of the two experts in our country, “Those with dry scalps are generally more prone to dandruff. And due to the problem of dandruff, the problem of hair fall occurs. On the other hand, those with oily scalps can easily get rid of this problem if they clean their hair well.” Also, if the scalp is oily, regular shampooing will gradually get rid of this problem. However, those with dry skin need to be a little extra careful. Moreover, hair specialist Shivani De says, “In addition to internal care, taking care of hair externally also requires regular oil massages, proper combing and use of protein packs.” He also said that using natural protein packs twice a week helps to eliminate various hair problems.

In this research paper, we will be using machine learning

to find the various reasons why falling hair falls from our heads. Our main reason for this paper is how many ways we lose our hair. What kind of problem does our hair fall at a young age? In the following other papers, they used Machine Learning Model SVM, KNN, Logistic Regression, Random Forest. In our paper, we also use these algorithms and besides this algorithm, we use XGBoost for better accuracy.

The rest of the paper is structured as follows. Here are three segments. In segment-2, we discuss the related work described followed by research. In segment-3, Discuss the Research Methodology. In segment-4, Result Discussion and analysis of our results. In segment-5, Machine Learning Model Analysis. In segment-6, Result and Discussion. Finally, We conclude this paper.

## II. LITERATURE REVIEW

Based on our research topic (Hair Fall) we picked some papers as an initial study. The study of research outcomes would help to gain the end. From this review portion, we will know other's opinions about Hair Fall.

Kim, M., Kang et al.[1] measured hair density using Deep Neural Network. Using that method, they were able to measure hair density, measure hair fall intensity, count the number of donor hairs, etc. They worked on 4492 datasets for training and evaluation. Performance EfficientDet, YOLOv4, and DetectorRS algorithms were experimentally used for object detection for comparison. YOLOv4 gave the best performance with an average accuracy of 58.67%. Almohanna et al.[2] conducted extensive research on Alopecia Areata Diagnosis for Hair Loss Related Autoimmune Disorder in Women. The doctor is the one who determines whether or not a patient has a sickness or illness. Alopecia areata is a chronic disorder that results in hair loss in the afflicted region. They used models to extract features in the work, with 70% of the database used for training and 30% used for testing, resulting in a 7:3 split. As a consequence, use techniques such as ANN, SVM, and Logistic regression to achieve maximum accuracy of 98.3%. Benhabiles et al.[3] used Deep Learning to determine the amounts of hair loss in guys using face photos. Furthermore, despite the limited visibility of hairs in such photographs, a matching approach for automatically classifying face images in relation to pattern classification tables of male baldness from the medical field is provided. The results of the experiments demonstrate the potential and efficiency for medical, security, and commercial applications. Average accuracies obtained on the test set classification utilizing three CNN-based architectures over baseline and enhanced training sets were compared. Although the DEX-IMDB-WIKI achieved the highest average accuracy of 85.5 percent. Sepenu, A.K et al.[4] In order to maximize the effectiveness of current sales and customer segmentation models and strategically plan their marketing and sales initiatives, a professional hair product manufacturer has attempted to apply machine learning models such as clustering and regression. The proposed clustering and regression models showed comparisons between logistic regression and random forest classifier models using only internal company-provided

data. Using random forest in this comparison gave the best performance with accuracy at 98.3%. Shabnam Sayyad et al. [5] was analyzed Alopecia areata related with hair loss . Alopecia areata causes baldness in humans, their study discussed ML and deep learning can help diagnose this disease. Choudhary Sobhan Shakeel et al. [6] was studied to identify hair loss caused by the skin disease Alopecia areata. Hair image was trained for this study.They used SVM and KNN models and SVM gave 91.4% accuracy while KNN gave 88.9% accuracy. ALEnezi et al. [7] proposed pre-trained CNN for image processing to detect skin disease. Then they used multiclass SVM for classified features. Their system detected 3 skin diseases with 100% accuracy. Allugunti et al. [8] classified skin disease as melanoma. They used DL and ML algorithms such as DT, RF, GBT, and CNN. Of all algorithms, CNN gave the highest accuracy of 88.83%. Bhadula et al. [9] detected skin disease using ML five algorithms such as Random forest, naive bayes, logistic regression, kernel SVM, and CNN. CNN was giving best accuracy of training 99.05% and testing 96%. Kapoor et al. [10] were classified using the suggested systems feedforward artificial neural network and backpropagation method for patients with and without alopecia. According to their experimental findings, the accuracy is 91%. Weerasinghe et al. [11] proposed a new face shape classification method for choosing the best haircut or hairstyle. To detect, they employed ML classifications like Naïve Bayes. They used five thousand data and both the face shape classification model and the hair recommendation model are 83% and 91% accurate, respectively. Yacoob et al. [12] purposed of comparing the hair appearance of various persons, they created computational models. Results were provided together with a description of an automatic hair-detecting system. they compared eigenface-based recognition and a combined eigenface-hair-based identification for the identification of individuals using hair features.

Our research has revealed a lot of novelties .our dataset is collected in an excel sheet through a survey so our dataset is authentic. Also, we have identified some specific causes of hair loss like blood group, stress, allergies, etc. In previous works that did not use XGBOOST we proposed a new algorithm XGBOOST and got good results.

## III. RESEARCH METHOD

This research paper displays the participation of Bangladeshi people in hair fall using machine learning algorithms. To accomplish the research objectives, we followed the following steps.

### A. The significance of hair fall survey:

The fundamental purpose of our survey analysis is to identify the most significant problems to fall hair. Most of the audience is between the ages of 18-24. Those who are studying university. This time period is the most energetic time for both male and female everyone in life. At that time we are facing hair fall. This survey finds lots of reasons for falling hair. If any of the parents or other members of the family have

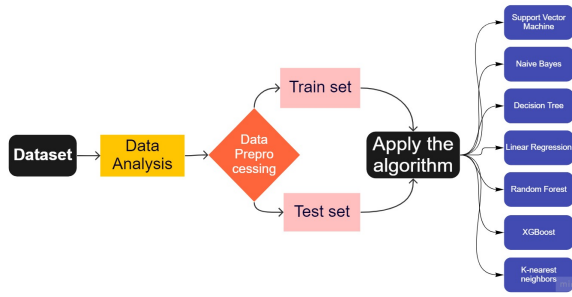


Fig. 1. Stages of the proposed procedure

a problem with hair loss, the chances of a boy or girl for hair loss increase. Also, problems like allergies, anxiety, stress and dandruff can cause hair loss.

### B. Data Collection:

Our research aims to find the cause of excess hair loss using machine learning algorithms. We create a guide on how to collect data through surveys and find out what causes excess hair loss.

- First, we chose hair loss symptoms to create the questionnaire.
- We create an online form and make some questionnaires.
- Through this form, we collect data on Bangladeshi men and women aged 14-50.
- After collecting the data we test and preprocess the dataset using machine learning.
- We have analyzed the causes of excess hair loss.

After collecting all the samples, our sample number is 610. As we know, data is the most important component of machine learning and generating a dataset is the most important research study activity. Pre-processing of the dataset includes assessing the quality of the data, discarding all null values, cleaning the data, properly labeling the datasets as they are created, transforming the data and reducing the data to make the dataset suitable for building and training machine learning models. Apply the algorithm by dividing the dataset into train and test sets.

### C. Data Analysis:

A detailed discussion about any data is called data analysis. Our Dataset consists of 610 opinions, the number of male responses has 269 and female responses has 341. The dataset has a total of 16 questions. There are five parts to complete our survey research paper data analysis.



Fig. 2. Step of hair fall survey data analysis process

### D. Description of the survey form:

We have created a Google Form based on all these issues. Since this survey is done in the context of Bangladesh, we have made our questionnaire in both Bengali and English languages so that all types of people can understand it easily. All the factors we have considered while developing the questionnaire for the research. Among them, the important questions are:

TABLE I  
QUERY DESCRIPTION OF OUR DATASET

Features	Description
Age	Age represents how many people of any age shared their opinions with us.
Gender	We decided to collect data from Male and Female.
Educational Institution	What kind of educational institution are they representing? School, college, university, and others. Others mean Job holders, Housewife etc.
Location	We want to know which districts people are sharing their thoughts
Excess hair fall	This part has two questions- 1.Yes 2. No
How long have you been suffering from hair loss	To find out how long people with receding hair have been suffering from this problem.
Does anyone in your family have such a problem?	We are trying to know if anyone facing hair fall problems in Hereditary.
Dandruff problem	Does dandruff cause hair loss? Here we know how many people have dandruff.
Allergy problems	If someone has an allergy problem, what percentage of them suffer from hair loss problem?
Skin type	To know whether skin type is associated with hair loss or not, let's find out which skin type people have more hair.
Blood group	Through the blood group, it will be known which blood group people have hair fall.
Stress	Now-a-days people lead very busy lives. As a result, the stress increases and it is known how many people lose their hair due to this stress

## IV. RESULT DISCUSSION AND ANALYSIS

18-24 years old participated the most in our survey. Their number was 296 out of a total of 610. That is, 57% of the participants are in the age range of 18-24 years. So we are clear that our main participants are men and women aged 18-24. Also, we have two different age groups. 13-17 and 25-34. The total percentages are 0.3% and 42.3% respectively.

we collected data from both males and females. Among them, the largest number of participants are women. Females are 55.9% while males are 44.1%.

Through this, we know the background of the participants. Here we gave school, college, university and others options, others means jobholders, unemployment, serviceholder etc. 55.7% of the participants are university students, others are 43.1% and college students 0.7%.

We collected data from people from eight districts of Bangladesh. While the highest participant in Dhaka is 25.9%, the participant in Sylhet is the second highest 16.4%, and the participants in Chittagong (11.5%) and Rangpur (11.8%)

are almost equal. The least participants are Barisal 8.2%, Moreover Mymensingh, Rajshahi, and Khulna are 8.9%, 8.7% and 8.7% respectively.

Most of the participants in our survey are facing hair fall problems. The rate is 58.0%. The affected rate is 353 people of a total of 610. Besides 42% are not suffering any hair falling.

This question and some options are given to find out how long the hair loss sufferers have been suffering from this problem. Hair loss for more than 12 months is 26.9%, 1 or 6 months is 18%, and 7 or 12 months is 15.1%. Only 0.5% of them had normal hair loss. Also, 0.5% of people are suffering for 3 years.

Hair loss is often caused by genetics. That's why we wanted to know Mainly whether anyone in their family has a hair loss problem. Where the rate of hair loss in the family is very high at 55.6% and 44.4% of the participants have no hair loss problem in the family.

Dandruff is one of the causes of hair fall due to various reasons in our daily life. Here we see that 57% of people are suffering from dandruff problems while 43% of people are not having dandruff problems.

Currently, due to climate change, the temperature is increasing and we are having different types of skin problems, so we have seen whether anyone has skin problems or not, and it is seen whether skin problems affect hair loss or not. Here we see that 52.8% of people's skin There are problems or allergy problems, and 47.2% of the participants do not have allergy problems.

It is important to keep our mental health good. But our daily life has become stressful due to which there are many problems but here it is seen whether those of us who have hair fall are suffering from stress problems or not. 57.5% of participants lead a stressful life and 42.5% of participants lead a stress-free life.

It can be seen that 35.2% of the participants had oily skin type, 26.6% had combination skin type, 19.2% had dry skin type and 19% had normal skin type. Many times we use different products without understanding the skin type which results in our hair falling.

In this way, we can know which blood group people have more hair loss and we can understand whether the blood group is related to hair loss or not. Among the participants, the B+ blood group is 27.2%, AB+, and A+ blood groups are 18.2%, the O- is 17.7% and the least participant is AB- blood group 1.1%.

Previously we are showing 58% of people are suffering from hair falls. Only 21.6% of them use medicine for their hair care. We can see that 78.4% of people are not taking care of their hair. Very few people are serious about their hair loss.

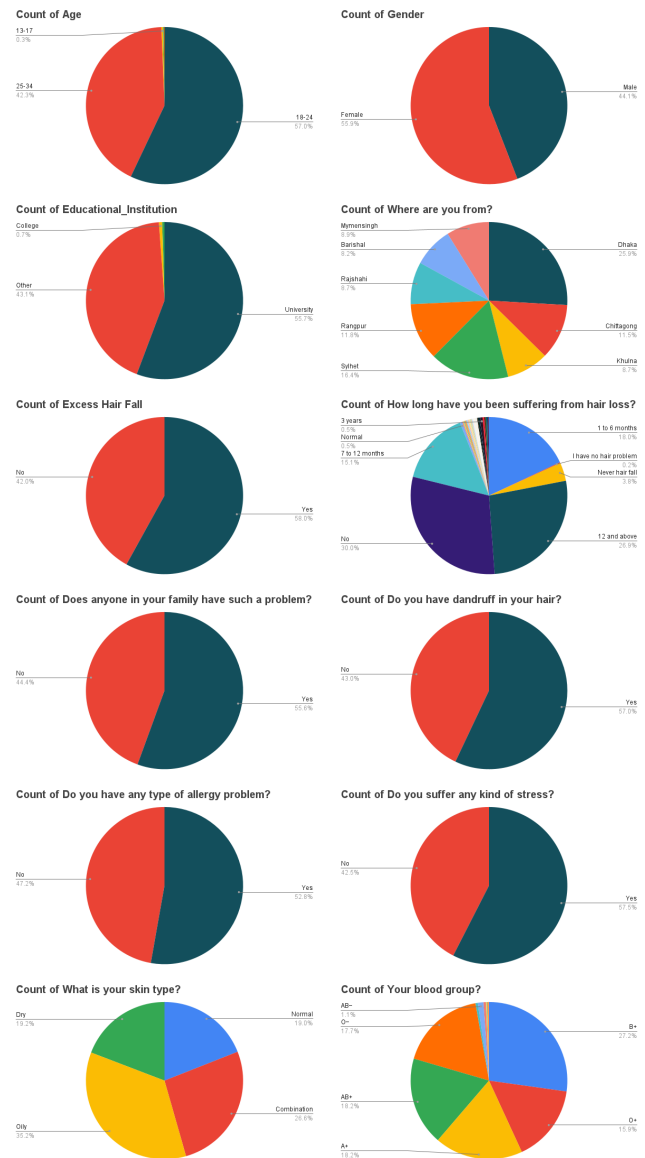
Hair is one of the symbols of beauty. In our survey, we can see most of the people are suffering from hair fall, and most of the participants are female. The Interesting fact is only 51.3% of people take care regularly their hair. Although the rate is not bad. Besides 48.7% are not taken care of regularly.

In previous Questions, 51.3% of people take care regularly their hair. They use shampoo, oil, Aloe Vera, Coconut Oil, etc.

Most people are not only using one product. They use at a time multiple products. Someone uses Shampoo, Oil, and aloe vera, someone oil, and Shampoo and someone only has coconut oil. Only Shampoo users are 11.5%. Also, some people are not serious about their hair. 43.3% of people are not using anything in their hair.

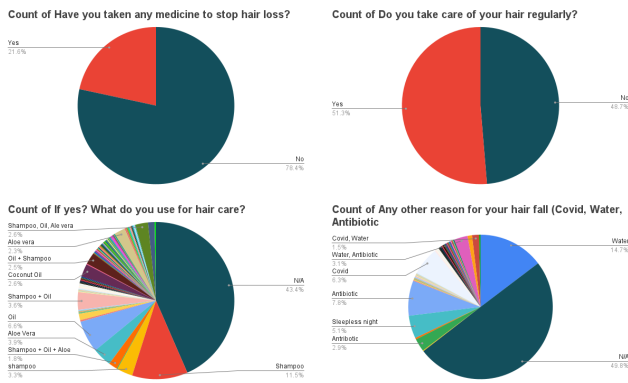
The participants are finding some reasons for falling hair. Covid-19 is one of the reasons. Last two years we went through a pandemic. As a result of covid-19 infection, various diseases have appeared in our bodies. Besides, Water, Sleepless nights, and taking antibiotics are also one of the reasons.

### V. SAMPLE DATASET



### VI. MACHINE LEARNING MODEL ANALYSIS

Since all of the data is 100% accurate then we marked them for further analysis. It is split into train and test sets



and then prepped for model input. With machine learning, a user provides lots of algorithm with a large amount of data and the computer will analyze it and make a decision based on the data provided. Here we used 7 important algorithms which are Support Vector Machine, Decision Tree, Logistic Regression, Naive Bayes, Random Forest, XGBoost Classifier and K-nearest neighbors algorithm. Now we will discuss the algorithms we have used.

**Support Vector Machine (SVM):** Finding the best line in two dimensions or the best hyperplane in more than two dimensions uses SVM models to help us classify the space.

**Decision Tree (DT):** By understanding simple decision rules derived from prior data, a decision tree is used to build a training model that may be used to predict the class or value of the target variable.

**Random Forest:** In order to reduce the variance, numerous deep decision trees trained on various subsets of the same training data are averaged using random forests. This significantly improves the performance of the final model at the cost of a slight increase in bias and a slight reduction in interpretability.

**XGBoost Classifier:** XGBoost full form is Extreme Gradient Boosting. It is a scalable and distributed gradient boosted decision tree Machine Learning library.

**K-nearest neighbors (KNN):** The KNN algorithm, which classifies or forecasts the clustering of individual data elements using proximity. It is a non-parametric and supervised learning classifier.

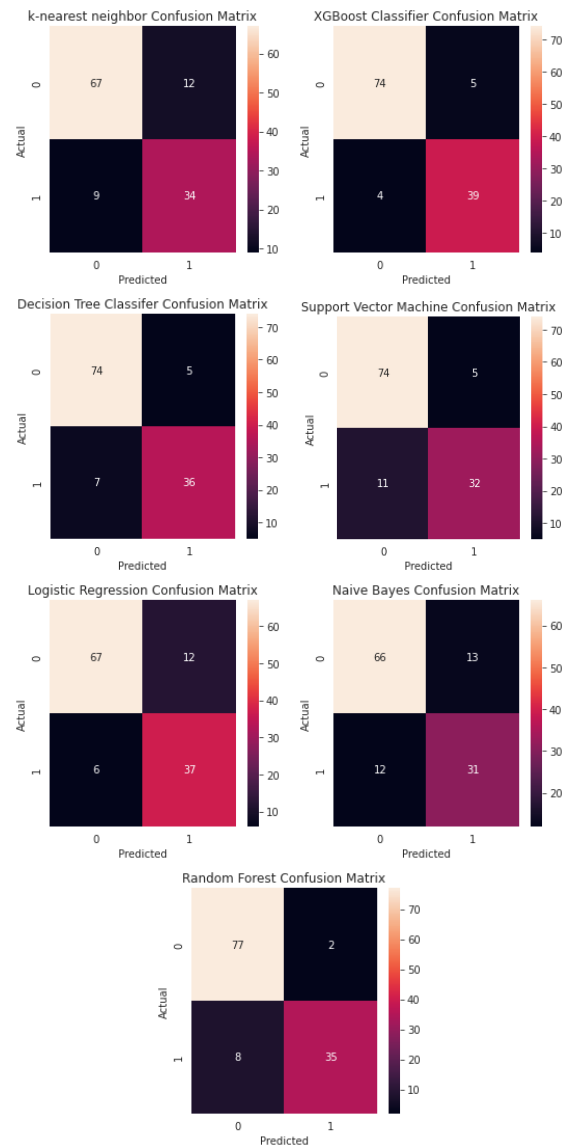
**Naive Bayes:** Naive Bayes classifier is a simple probabilistic classifier established on Bayes theorem but with strong independence assumptions.

**Logistic Regression:** In contrast to multiple linear regression, which analyzes predictors of outcomes with a continuously distributed distribution, multiple logistic regression assesses predictors of outcomes with a binary distribution, or outcomes that either occurred or did not.

A. Confusion Matrix

The confusion matrix considers a binary process of classification. The resulting table consists of two rows and two columns, filled with four values - true positive, false positive, true negative, and false negative. In the confusion matrix, a

true positive occurs where the observation is positive with a positive predictor, and a false positive occurs where the observation is negative, with a positive predictor. A true negative occurs where the observation is negative with a negative prediction, and a false negative indicates a positive observation with a negative prediction.



VII. RESULTS AND DISCUSSION

We find lots of reasons for falling hair. We are collecting data for three months. Our dataset has 610 data. More than half of the total dataset is female. The percentage rate is 55.9%. Most of them are university students. And Their age range is 18-24. We identified four main reasons for falling hair: stress, allergy, dandruff and family history. Besides, 51.3% of People take care of their hair regularly. They use Shampoo, Aloe vera and Oil. 48.7% are not caring about their hair regularly. Some people take medicine to stop hair falls. The percentage rate is not so high. The rate is 21.6%. Most people are not taking any

kind of medicine. Also, we identify, the B+ blood group and Skin type oily people lose their hair. The reason for the hair fall is that people share their opinion: Covid-19, Water, Antriboitic, and Sleepless night. Overall 58.0% of people face hair falling. And 42.0% are not facing any hair fall problems. After the survey, we are trying to analyze the survey using machine learning algorithms. The accuracy rate is 92.62% which is an XGBoost Classifier. Respectively Decision Tree - 90.16%, Random Forest - 91.80%, Support Vector Machine - 86.89%, Logistic Regression - 85.25%, Naive Bayes - 79.51% and K-nearest neighbor - 83%.

#### A. Accuracy and F1-score Table:

This Accuracy table shows that six algorithms are used. XGBOOST gave the best accuracy for this dataset. Besides, Random Forest & KNN also gave good accuracy. SVM & logistic regression gave the same accuracy. The decision tree gave the lowest accuracy.

TABLE II  
ACCURACY TABLE AND F1-SCORE

Model	F1-score	Accuracy
Decision Tree	0.90	90.16%
Random Forest	0.92	91.80%
Support Vector Machine	0.87	86.89%
XGBoost Classifier	0.93	92.62%
Logistic Regression	0.85	85.25%
Naive Bayes	0.80	79.51%
K-nearest Neighbor	0.83	83%

#### B. Performance Comparison

Here we take 80% data for training & 20% data for testing. Then after training all the models we see that we got the highest accuracy of 92.62% from XGBOOST model & second highest accuracy 91.80% from Random Forest model. Decision Tree got the third highest accuracy of 90.16%. SVM & Logistic Regression were almost similar 86.89% and 85.25% accuracy. K-nearest gives quite better accuracy which is 83% and the lowest accuracy obtained from the naive bayes is 79.51%. From this, it can be seen that the XGBOOST model is the best model for this dataset.

### VIII. CONCLUSION AND FUTURE WORK

We conducted a survey on them to find out what causes hair loss and balding at a young age. Also, we use machine learning algorithms. With a 92.62 percent overall accuracy rate, our suggested technique has excellent anticipated accuracy at each stage. In the next step, we will use technologies like deep learning, image processing, etc. to investigate the cause of hair loss. In the future, We will collect images of men and women and use computer vision technology. Besides, we will collect men's and women's weights and heights and calculate their Body mass index (BMI).

### REFERENCES

- [1] Kim, M., Kang, S. and Lee, B.D., 2022. Evaluation of Automated Measurement of Hair Density Using Deep Neural Networks. *Sensors*, 22(2), p.650.
- [2] Almohanna, H.M., Ahmed, A.A., Tsatalis, J.P. and Tosti, A., 2019. The role of vitamins and minerals in hair loss: a review. *Dermatology and therapy*, 9(1), pp.51-70.
- [3] Benhabiles, H., Hammoudi, K., Yang, Z., Windal, F., Melkemi, M., Dornaika, F. and Arganda-Carreras, I., 2019, November. Deep learning based detection of hair loss levels from facial images. In 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA) (pp. 1-6). IEEE.
- [4] Sepenu, A.K. and Eliassen, L., 2022. A Machine Learning Approach to Revenue Generation within the Professional Hair Care Industry. *SMU Data Science Review*, 6(1), p.6.
- [5] Sayyad, S. and Midhunchakkaravarthy, D., Deep Study on Alopecia Areata Diagnosis for Hair Loss Related Autoimmune Disorder Problem Using Machine Learning Techniques. Available at SSRN 4054490.
- [6] Shakeel, C.S., Khan, S.J., Chaudhry, B., Aijaz, S.F. and Hassan, U., 2021. Classification framework for healthy hairs and alopecia areata: a machine learning (ml) approach. *Computational and mathematical methods in medicine*, 2021.
- [7] ALEnezi, N.S.A., 2019. A method of skin disease detection using image processing and machine learning. *Procedia Computer Science*, 163, pp.85-92.
- [8] Allugunti, V.R., 2022. A machine learning model for skin disease classification using convolution neural network. *International Journal of Computing, Programming and Database Management*, 3(1), pp.141-147.
- [9] Bhadula, S., Sharma, S., Juyal, P. and Kulshrestha, C., 2019. Machine learning algorithms based skin disease detection. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(2), pp.4044-4049.
- [10] Kapoor, I. and Mishra, A., 2018. Automated classification method for early diagnosis of alopecia using machine learning. *Procedia computer science*, 132, pp.437-443.
- [11] Weerasinghe, H. and Vidanagama, D., 2020, December. Machine learning approach for hairstyle recommendation. In 2020 5th International Conference on Information Technology Research (ICITR) (pp. 1-4). IEEE.
- [12] Yacoob, Y. and Davis, L.S., 2006. Detection and analysis of hair. *IEEE transactions on pattern analysis and machine intelligence*, 28(7), pp.1164-1169.