International Journal of Engineering Approaches

Araştırma Makalesi / Research Article

Investigation of the effect of various factors on hair loss using machine learning techniques

Çeşitli faktörlerin saç dökülmesine etkisinin makine öğrenmesi teknikleriyle incelenmesi

Murat Aslan[1]*, Berkant Baş[1], Bekir Parlak[1]

[1] Amasya University, Department of Computer Engineering, Amasya, Turkey

* *Sorumlu Yazar / Corresponding Author:* bekir.parlak@amasya.edu.tr

Makale Bilgileri / Article Info

Keywords
Algorithm
Hair Loss
Engineering
Original studies
Machine Learning
Data Mining

Anahtar Kelimeler
Algoritma
Saç Dökülmesi
Mühendislik
Özgün çalışmalar
Makine Öğrenimi
Veri Madenciliği

Makale tarihçesi / Article history
Geliş / Received: 24.04.2025
Düzeltme / Revised: 15.05.2025
Kabul / Accepted: 22.06.2025

Abstract

Nowadays, hair loss has become a big problem for people in terms of psychology, aesthetics and many other aspects. Anxiety, stress, irregular nutrition, genetic and environmental factors are among the main causes of hair loss. This study was carried out to determine the various factors affecting hair loss and to test the suitability of machine learning and data mining methods in this study process. Analyses were made with many machine learning algorithms using different data sets. According to the results, while coffee consumption, which is one of the factors that has the most effect on hair loss in the first data set, was seen to affect hair loss by 95%, the factors in the second data set were seen to have less effect on hair loss compared to the factors in the first data set. These results show that we can use machine learning algorithms as an effective tool in the process of better understanding the hair loss problem and early diagnosis.

Öz

Günümüzde saç dökülmesi, insanlar için psikolojik, estetik ve birçok açıdan büyük bir sorun haline gelmiştir. Kaygı, stres, düzensiz beslenme, genetik ve çevresel faktörler saç dökülmesinin temel nedenleri arasında yer almaktadırlar. Bu çalışma da saç dökülmesini etkileyen çeşitli faktörleri belirlemek ve bu çalışma sürecinde makine öğrenimi ve veri madenciliği yöntemlerinin uygunluğunu test etmek amacıyla gerçekleştirilmiştir. Farklı veri setleri kullanılarak birçok makine öğrenimi algoritmaları ile analizler yapılmıştır. Sonuçlara göre ilk veri setinde saç dökülmesi üzerinde en çok etkiye sahip etkenlerden kahve tüketimi %95 oranında etkilediği görülürken, ikinci veri setindeki etkenler ilk veri setindeki etkenlere oranla saç dökülmesine etkisi daha yetersiz olduğu görülmüştür. Bu sonuçlar saç dökülmesi sıkıntısının daha iyi idrak edilmesi ve erken teşhis sürecinde makine öğrenimi algoritmalarının etkin bir araç olarak kullanabileceğimizi göstermektedir.

1. Introduction

Hair is considered an important component of an individual's overall appearance [1]. Hair loss is a condition that most individuals can experience and occurs because of genetic and environmental factors. Hair loss is not life-threatening, but it is distressing and significantly impacts the patient's quality of life [2]. For most people, hair loss brings with it a loss of self-esteem, negative effects on social life, and increased feelings of depression [3]. Health factors such as protein and keratin amounts, which are important for the structure of the hair, the number of micronutrients and body water content have a significant effect on hair loss. The human scalp contains approximately 100,000 hair follicles, 90% of which require essential elements such as proteins, vitamins and minerals to produce healthy hair [4]. In addition, factors such as stress levels are often considered as causes of hair loss.

This article aims to examine the factors affecting hair loss through two data sets. The aim is to determine the main factors that have a significant impact on hair loss based on the data. It has been some suggestions and comments to reduce hair loss based on these results.

1.1. Scientific studies on hair loss

Artificial neural networks can be used to indicate the frequency of hair loss. Some of the reasons affecting hair loss, the parameters evaluated by the doctor and the hair specialist are used in neural networks [7]. Artificial Neural Networks are widely used for pattern recognition and system detection. Networks consist of nonlinear computational objects moving in parallel. The feature of ANN is that they can identify the environment and increase efficiency at the time of learning [8-10] Thanks to this efficiency, the methods used are used to detect hair loss. With the regression estimate and the percentage output it gives, it has been tried to make the work of experts easier

and to prevent a loss of time in interpreting the output of long-running analyses. Thanks to the classification methods, early diagnosis has been made easier for humanity by obtaining the results of whether it is present or not. Due to these limitations, advanced tools and techniques have been developed and tested using deep learning. Gender, age, zinc content, iron content, anemia and cosmetic use. These methods have been shown to be highly successful with a neural network built with the Levenberg-Marquardt algorithm [11]. When using Support Vector Machine (SVM), the separation of Gauss (DoG) filters is considered for feature determination [12]. Deep learning to classify amounts of baldness and hair loss in facial photographs [13] or other methods have been introduced, such as classifying four common scalp hair signs [14]. In a study [20], it is proposed a deep learning model capable of analyzing scalp disorders with high accuracy. Utilizing this model, researchers may anticipate scalp diseases, facilitating effective treatment through mobile phones by merely providing images of the conditions. Also, it [21] introduces an advanced healthcare platform for assessing the severity of six prevalent scalp hair disorders: dryness, oiliness, erythema, folliculitis, dandruff, and hair loss. To develop an appropriate scalp image classification model, we evaluated three deep learning architectures: ResNet-152, EfficientNet-B6, and ViT-B/16. Of the three evaluated deep learning models, the ViT-B/16 model had the highest classification performance, achieving an average accuracy of 78.31%. In the study [22], it employs deep learning algorithms on photos of hairy scalps. Issues related to a hairy scalp are typically detected by non-professionals at hair salons, who may offer advice to those experiencing these concerns. Moreover, several prevalent scalp issues exhibit similarities, leading to potential misdiagnoses by non-experts. Consequently, scalp issues have deteriorated. The study involved the implementation and comparison of the deep-learning technique, namely the ImageNet-VGG-f model Bag of Words (BOW), alongside machine-learning classifiers, as well as the histogram of oriented gradients (HOG) and pyramid histogram of oriented gradients (PHOG) with machine-learning classifiers.

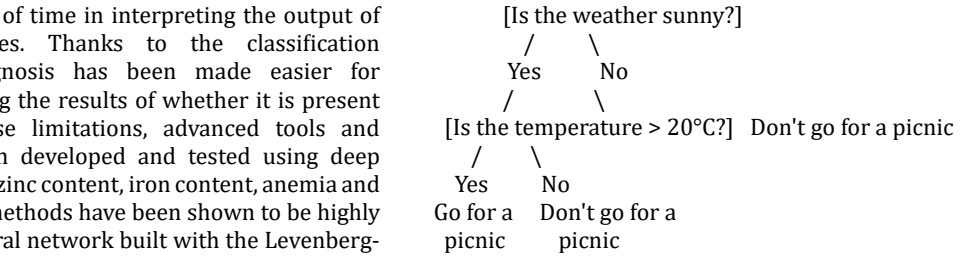
2. Methodologies

In this study, classification and regression algorithms were used. Classification algorithms from these algorithms; Decision Trees, Random Forest, Support Vector Machines (SVM) algorithms. Regression Algorithms, Random Forest, Decision Trees, XGBoost algorithms.

2.1 Classification

2.1.1. Decision tree

Decision trees are one of the most important methods for data mining; they provide us with ease of use. It is a supervised classification model that is generally used in many methods because it tries to eliminate uncertainties and is robust even in missing data [5]. Decision trees create a model that separates examples from a dataset based on a set of decision rules. These rules are organized into branches of the tree, and each branch divides and classifies the data according to a specific feature. Below is a decision tree model that shows whether to go on a picnic if the weather is sunny or not.



Explanation:

- First question: **Is the weather sunny?**
- If **Yes**, check the temperature.
- If it's **above 20°C** → ✅ Go for a picnic.
- If it's **not** → ❌ Don't go for a picnic.
- If **No**, directly → ❌ Don't go for a picnic.

2.1.2. Random forest

It is one of the supervised learning algorithms and is used to solve classification and regression problems. The Random Forest classifier is a meta-estimator that fits a series of Decision Tree Classifiers to various subsamples of the dataset and uses averaging to improve prediction accuracy and control overfitting [15].

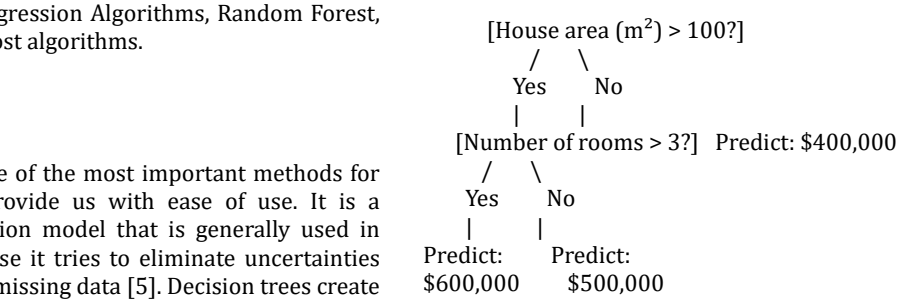
2.1.3. Support vector machines

It is one of the supervised learning algorithms and is used to solve classification and regression problems. It is an effective method especially for nonlinear classifications and is widely used in classification problems. The basic logic of the Support Vector Machine Algorithm, which is the Support Vector Machine (SVM), is to determine the best solution that can be divided into data lines [16].

2.2. Regression

2.2.1. Decision tree

A regression decision tree is an algorithm used to predict numerical values. It treats the data as a tree, divides it into branches, and calculates a numerical value estimate for each branch. This estimate is calculated by averaging the samples. It is often used in complex problems.



Explanation:

- **Root node:** Is the house larger than 100 m^2?
- If **yes**, check: Are there more than 3 rooms?
- At the **leaf nodes**, the tree gives a **numerical prediction** (house price).
- These predictions are usually the **average** of training examples that fall into that path.

2.2.2. Random forest

Random Forest Regression and Decision Tree Regression, which work in the background, try to explain the trained and tested data in detail [6]. These models work by learning patterns from the training data and dividing the data into smaller subsets based on the feature values. In Decision Tree Regression, a single tree iteratively splits the input region and places a fixed value at each leaf node. On the other hand, Random Forest Regression creates an ensemble of multiple decision trees, each trained on a random subset of the data and features and averages their outputs to reduce overfitting and increase prediction accuracy.

2.2.3. XGBoost

XGBoost is an optimized gradient boosting library that works with decision trees. Unlike random forest, when creating the final estimate in XGBoost, the linear sum of all trees is taken and the goal of each tree is to minimize the residual error of the previous trees [17]. In other words, each new tree is trained to predict the residuals of the combined prediction of the previous trees, which gradually increases the accuracy of the model with each iteration. This sequential learning strategy allows XGBoost to focus on data points that are harder to predict, effectively reducing bias and variance.

3. Datasets

Two different datasets were used in this study. These datasets were taken from the Kaggle website [18][19]. Our first dataset includes many features based on hair loss. These are hair loss, staying up late, pressure level, amount of coffee consumed, brain working hours, stress level, shampoo brand, swimming, hair washing, hair oiliness, dandruff and libido levels. In total, there are 401 data and 13 features in this dataset. Regression and classification estimates were made by taking these features into account. The other dataset also includes many features based on hair loss. These are total protein, total keratin, hair texture, vitamin, manganese, iron, calcium, body water content, stress level, liver data and hair loss. There are 430 data and 11 features in total. Regression and classification estimates were made in light of these features. Although some data features are similar, most of them differ. The causes of hair loss were examined in more detail by comparing them.

4. Experimental Study

Table 1. Values and ranking factors affecting hair loss

Feature	Importance
Coffee_consumed	0.240763
Libido	0.166285
Stay_up_late	0.088098
Day	0.080619
Brain_working_duration	0.071987
Hair_grease	0.070410
Stress_level	0.062443
Dandruff	0.057863
Pressure_level	0.042578
Month	0.042410
School_assesssment	0.031465
Hair_washing	0.026514
Shampoo_brand	0.009804
Swimming	0.007873
Year	0.000889

According to the Table 1, the primary influencing factor in the relevant model is coffee consumption (24.08%). The result here shows that coffee consumption is one of the strongest factors on hair loss. Secondly, there is the libido (16.63%) variable. These two factors play an important role in determining the prediction percentage of the model. The factors following this duo are, respectively, going to bed late (8.8%), day variable (8.06%), and brain working time (7.19%). These variables are called medium-level effects. It helps us make sense of the model, it is not as determinative as coffee consumption and libido. Other variables that follow the variables we mentioned, hair oil (7.04%), dandruff (5.76%), and pressure level (4.26%), have less effect compared to the others. The variables with the lowest effect are shampoo brand (0.90%), swimming (0.78%), and year (0.08%). These are the results we obtained by running the first data set we have with machine learning techniques on the necessary platform (Spyder).

Table 2. Classification accuracy rates with most influencing factor

ACCURACY	%85
F1-SCORE	%83.98
ROC-AUC	%85.03

According to the Table 2, coffee consumption and libido rates are the first data set are the highest and the study was examined according to these variables. Accuracy value is 85%, F1-Score value estimates this accuracy percentage as 83.98% with the balance of false positives and false negatives, ensuring that accuracy is guaranteed according to the F1-Score and Roc-Auc determines the discrimination power as 85.03%, indirectly affecting the accuracy percentage.

4.1.1. Classification results

Table 3. Classification results in dataset 1

Model	Accuracy	F1-Score	ROC-AUC
Random Forest	%92.5	%92.47	%98.69
Decision Tree	%95	%94.90	%90.42
SVM	%90	%90.04	%96.12

When we examine in the Table 3, Decision Tree has become the model that shows the most successful performance with a rate of (95%). When we examine the other models, it is seen that the intersections of Random Forest - ROC-AUC (98.69%) and Random Forest - F1 Score (92.47%) have a stronger overall performance. We see that the intersection of SVM - ROC-AUC (96.12%) has good discrimination power. However, the accuracy and F1-Score values are seen to be lower than the other models. The results show that different qualities should be given priority depending on the purpose of the application.

4.1.2. Regression results in dataset 1

Table 4. Regression results

Model	MAE	RMSE	R^2
Random Forest	%15.89	%36.27	%87.16
Decision Tree	%10	%35.36	%87.80
XGBoost	%13.74	%31.45	%90.35

When we take in the Table 4 as a reference, the XGBoost - R^2 intersection has reached the highest probability value in the table by reaching 90.35%. The XGBoost - RMSE intersection

(31.45%) has the lowest value. This shows that the model is more consistent compared to other models. Although the Decision tree model has a low MAE value (10%), we find that XGBoost is generally behind in terms of performance. When we examine the RF model, we see that it produces lower R^2 (87.16%) and higher error values. Based on these results, we can say that XGBoost is the most successful model.

4.2 Values and ranking of factors affecting hair loss according to the second data set related table

Table 5. Values of factors affecting hair loss

Feature	Importance
Vitamin	0.110190
Manganese	0.107312
Liver_data	0.107203
Calcium	0.102192
Total_protein	0.101526
Stress_level	0.098991
İron	0.098597
Hair_texture	0.096331
Body_water_content	0.089948
Total_keratine	0.0877711

According to the Table 5, the primary effect factor in the relevant model is vitamin (11.01%). The result obtained from this shows that vitamin is one of the strongest factors on the hair loss effect. The second variable is manganese (10.73%). These two factors play an important role in determining the prediction percentage of the model. The factors following this pair are late liver_data (10.72%), calcium (10.21%), total protein (10.15%), respectively. These variables mentioned are called medium-level effects. It helps us to make sense of the model, it is not as decisive as vitamin and manganese. According to the results of this research, vitamin and manganese rates are the highest and the study is examined according to these variables. These are the results we obtained by running the second data set we have with machine learning techniques on the necessary platform (Spyder).

Table 6. Classification accuracy rates with most influencing factor

ACCURACY	%23.26
F1-SCORE	%22.99
ROC-AUC	%55.88

According to the Table 6, accuracy value is 23.26%, and F1-Score value estimates this accuracy percentage as 22.99% with the balance of false positives and false negatives, allowing the accuracy to be guaranteed according to F1-Score, and Roc-Auc indirectly affects the accuracy percentage by determining the discrimination power at 55.88% of the accuracy.

4.2.1. Classification results in dataset 2

Table 7. Classification results

Model	Accuracy	F1-Score	ROC-AUC
Random Forest	%17.44	%17.39	%54.04
Decision Tree	%18.6	%18.86	%51.46
SVM	%10.47	%7.98	%50.84

When we examine the Table in 7, Decision Tree (18.6%) is the model that shows the most successful performance. When we examine the other models, it is seen that the intersections of Random Forest - ROC-AUC (54.04%) and Random Forest - F1 Score (17.39%) exhibited a stronger overall performance. We see that the intersection of SVM - ROC-AUC (50.84%) has a good discrimination power. However, the accuracy and F1-Score values of SVM are seen to be lower than the other models. The results show that different qualities should be given priority depending on the purpose of the application.

4.2.2. Regression results in dataset 2

Table 8. Regression results

Model	Accuracy	F1-Score	ROC-AUC
Random Forest	1.5643	1.8023	-0.0533
Decision Tree	2.0349	2.5404	-1.0926
XGBoost	1.6157	1.9293	-0.2070

When we take the Table in 8 as reference, the Decision Tree - RMSE intersection has reached the highest value in the table by reaching the rate of 2.5404. The Decision Tree - R^2 intersection has the lowest value of -1.0926. This shows that the model is more consistent compared to other models. We see that the RF model has a low MAE value (1.5643). When we continue with the RF model, it is seen that there is a higher R^2 value. When this table is taken as basis, it is seen that the most successful model is RF. It has the lowest MAE and RMSE values compared to other models. This makes this model the most successful.

4.3. Comparison of two experiments

The first data set showed better and higher performance in classification and regression cases, especially the Decision Tree model saw 95% accuracy. In addition, the most affecting factor, "coffee_consumed", provides a strong classification accuracy on its own. On the other hand, the second data set shows low performance in all models, and accordingly, we comment that it does not produce impressive results. In the second data set, we find the best classification accuracy (18.6%) in the Decision Tree. When we look at the first data set, a strong regression model stands out (XGBoost, R^2=%90). When we examine the second data set, we see that the R^2 values of the models are negative, and it is seen that they do not produce sufficient results in terms of explanation.

If we look at the result, the first data set gives more meaningful and more accurate results, has a stronger structure; while the second data set is weak in terms of the qualities we mentioned and the model.

Certain metrics were used while making these comparisons. These metrics are ACCURACY, F1-SCORE, ROC-AUC, MAE, RMSE and R^2.

Classification Metrics:

Accuracy: The ratio of the number of correctly predicted examples by the model to the total number of examples. It is useful if the data is balanced, but can be misleading in unbalanced data sets.

F1-Score: A balanced average of Precision and Recall values. It is a more reliable success measure, especially in unbalanced class distributions.

ROC-AUC: Shows how well the model distinguishes between classes. The closer the AUC value is to 1, the better the classification success of the model.

Regression Metrics:

MAE (Mean Absolute Error): It is the average of the absolute values of the differences between the prediction and the true value. It is easy to interpret.

RMSE (Root Mean Square Error): The errors are squared, averaged, and then the square root is taken. It is more sensitive to large errors.

R^2 (Coefficient of Determination): Shows how much the model explains the change in the data. The closer it is to 1, the more successful the model becomes.

To provide a clearer comparative overview for the reader, the performance metrics of all applied models across both datasets are summarized in Table 9 and Table 10 below. The best-performing metrics for each evaluation category are emphasized in bold for visual clarity.

Table 9. Accuracy, F1-Score, and ROC-AUC comparison of all models on both datasets

Dataset	Model	Accuracy (%)	F1-Score (%)	ROC-AUC (%)
1	**Decision Tree**	**95.00**	**94.90**	90.42
1	Random Forest	92.50	92.47	**98.69**
1	XGBoost	-	-	-
2	**Decision Tree**	**18.60**	**18.86**	51.46
2	Random Forest	17.44	17.39	**54.04**
2	SVM	10.47	7.98	50.84

Table 10. MAE, RMSE, and R^2 results of regression models applied on both datasets

Dataset	Model	MAE	RMSE	R^2
1	**Decision Tree**	**10.00**	35.36	87.80
1	Random Forest	15.89	36.27	87.16
1	XGBoost	13.74	**31.45**	**90.35**
2	**Decision Tree**	2.03	**2.54**	**–1.09**
2	Random Forest	**1.56**	1.80	–0.05
2	SVM	-	-	-

4.4. Performance metrics and their purposes

To comprehensively assess the performance of the machine learning models employed in this study, a variety of evaluation metrics were utilized. These include classification metrics—Accuracy, F1-score, and ROC-AUC—and regression metrics—MAE, RMSE, and R^2. Each of these serves a distinct analytical function in model validation and is well-supported in the machine learning literature [23–27].

Classification Metrics:

- Accuracy: Represents the proportion of correctly classified instances among all instances. While useful for balanced datasets, it may offer misleading insights in imbalanced scenarios [23].
- F1-score: The harmonic mean of precision and recall. Particularly effective in cases of class imbalance, where it provides a more holistic performance measure than accuracy alone [24].
- ROC-AUC (Receiver Operating Characteristic - Area Under Curve): Reflects the model's ability to distinguish between classes. A value closer to 1.0 indicates a higher discriminative capability. It is especially useful when evaluating model behavior across different thresholds [25].

Regression Metrics:

- MAE (Mean Absolute Error): The average of the absolute differences between predicted and actual values. It is easy to interpret and less sensitive to outliers [26].
- RMSE (Root Mean Square Error): The square root of the average of squared errors. It penalizes larger errors more heavily and is suitable for applications where large deviations are undesirable [26].
- R^2 (Coefficient of Determination): Indicates the proportion of variance in the dependent variable explained by the model. A value close to 1 signifies high explanatory power, whereas negative values imply the model performs worse than a naive mean-based predictor [27].

The combined use of these metrics ensures a thorough and multidimensional evaluation of model performance, covering not only correctness and discrimination but also prediction stability and variance explanation.

5. Discussion

This study involves analyzing the factors affecting hair loss using machine learning methods using two different data sets. The first data set showed high performance in both classification and regression models, and as seen in Table 3, the Decision Tree model stood out with 95% accuracy. In addition, coffee consumption ("coffee_consumed") proved to be an important factor by providing strong classification accuracy on its own.

On the other hand, the second data set showed low performance in all models and could not produce impressive results. The best classification accuracy in this data set (18.6%) was obtained again with the Decision Tree model, as seen in Table 7. When the regression analysis was examined, the XG Boost model showed the strongest performance with 90% R^2 value in the first data set. However, the negative R^2 values of all models in the second data set reveal that the explanatory power of the data set is weak.

Additionally, other regression metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were also taken into account to assess prediction stability and consistency. In the first dataset, the XGBoost model produced the lowest RMSE (31.45), indicating the lowest overall deviation between predicted and actual values, especially in penalizing large errors. On the other hand, the Decision Tree model achieved the lowest MAE (10.00), reflecting smaller average prediction errors across instances. The closeness between MAE and RMSE values further supports the stability of these models. Conversely, in the second dataset, all models demonstrated significantly higher MAE and RMSE values, which—when considered alongside the negative R^2 values—confirms their limited predictive capability and suggests a weak structural relationship between input features and the hair loss outcome.

As a result, the first data set was both significant and had a stronger structure, while the second data set was insufficient in these qualities. The analysis shows that the effect of coffee consumption on hair loss is great. These findings suggest that machine learning and data mining are powerful tools for early detection and awareness of hair loss.

In addition to the accuracy metric, which reflects the overall proportion of correct predictions, further evaluation was conducted using complementary classification metrics such as F1-score and ROC-AUC. The F1-score, defined as the harmonic mean of precision and recall, is particularly relevant when class imbalance is present, as it accounts for both false positives and false negatives.

In this study, the F1-score values were closely aligned with the accuracy results (e.g., 94.90% F1-score vs. 95% accuracy in the Decision Tree model), which confirms that the classifier maintained a balanced predictive performance across classes. Furthermore, ROC-AUC scores, which quantify a model's discriminative power across all classification thresholds, were notably high for Random Forest (98.69%) and SVM (96.12%), indicating strong class separation capabilities. These findings underscore the robustness of the models not only in overall correctness but also in handling class-level distinctions effectively.

6. Results

In this study, the effects of two data sets with different characteristics on hair loss were evaluated using machine learning methods. As a result of the analysis, it was shown which factor played an important role in which data set. It is tested the effects of the factors on hair loss with classification and regression estimates, evaluated their reliability according to the F1 scores and showed their accuracy rates, and when it is compared the two data sets, it is observed that coffee consumption had a very large effect on hair loss, as can be seen in the experiments. As a result of these observations, machine learning and data mining showed us that it is an important and powerful way for early diagnosis and awareness.

Conflict of Interest Statement: The authors declare no conflicts of interest.

Funding Information: This study was not supported by any funding.

Author Contributions: The authors confirm their responsibilities for the design of the study, data collection, analysis and interpretation of the results, and preparation of the manuscript.

Data Availability Statement: The data generated and/or analyzed during this study are not publicly available but can be provided by the corresponding author upon reasonable request.

References

[1]. Rushton, D. H., Norris, M. J., Dover, R., and Busuttil, N. (2002) *Causes of hair loss and the developments in hair rejuvenation,* International journal of cosmetic science, 24(1): 17-23.

[2]. Phillips, T. G., Slomiany, W. P., and Allison, R. (2017) *Hair loss: common causes and treatment,* American family physician, 96(6): 371-378.

[3]. Wells PA, Willmoth T, Russell RJ. (1995) *Talih mi lehinde? Kel mi? Erkeklerde saç dökülmesinin psikolojik bağlantıları.* Br J Psikoloji. 86:337–44.

[4]. Shapiro J. (2007) *Clinical practice Hair loss in women,* N Engl J Med. 357(16):1620–30.

[5]. Hastie TJ, Tibshirani, RJ, Friedman JH. (2009) *The Elements of Statistical Learning: Data Mining,* Inference and Prediction. Second Edition. Springer.

[6]. Nuray, S. E., Gençdal, H. B., and Arama, Z. A. (2021) *Zeminlerin kıvam ve kompaksiyon özelliklerinin tahmininde rastgele orman regresyonu yönteminin uygulanabilirliği,* Mühendislik Bilimleri ve Tasarım Dergisi, 9(1): 265-281.

[7]. Esfandiari, A., Kalantari, K. R., and Babaei, A. (2012) *Hair loss diagnosis using artificial neural networks,* International Journal of Computer Science Issues (IJCSI), 9(5): 174.

[8]. Haykin, S. (1999) *Neural networks: a comprehensive foundation*, Prentice hall.

[9]. Mandic D. P. and Chambers J. A., (2001) *Recurrent neural networks for prediction: Learning algorithms, architectures and stability,* Wiley.

[10]. Sureshkumar C. and Ravichandran T., *Character Recognition using RCS with Neural Network*.

[11]. Chang, W.-J.; Chen, L.-B.; Chen, M.-C.; Chiu, Y.-C.; Lin, J.-Y. (2020) *ScalpEye: A Deep Learning-Based Scalp Hair Inspection and Diagnosis System for Scalp Health,* IEEE Access, 8, 134826–134837.

[12]. Cho,T.S., Freeman, W.T., and Tsao, H. (2007) *A reliable skin mole localization scheme,* In Proceedings of the IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007.

[13]. Benhabiles, H., Hammoudi, K., Yang, Z., Windal, F., Melkemi, M., Dornaika, F., and Arganda-Carreras, I. (2019) *Deep Learning based Detection of Hair Loss Levels from Facial Images,* In Proceedings of the Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA), Istanbul, Turkey, 6–9 November 2019.

[14]. Chang, W.-J., Chen, L.-B., Chen, M.-C., Chiu, Y.-C., and Lin, J.-Y. (2020) *ScalpEye: A Deep Learning-Based Scalp Hair Inspection and Diagnosis System for Scalp Health,* IEEE Access 2020, 8, 134826–134837.

[15]. Veranyurt, Ü., Deveci, A., Esen, M. F., and Veranyurt, O. (2020). *Makine öğrenmesi teknikleriyle hastalık sınıflandırması: Random Forest, K-Nearest Neighbour Ve Adaboost Algoritmaları Uygulaması,* Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi, 6(2): 275-286.

[16]. Çiçek, A., and Arslan, Y. (2020) *Müşteri kayıp analizi için sınıflandırma algoritmalarının karşılaştırılması,* İleri Mühendislik Çalışmaları ve Teknolojileri Dergisi, 1(1): 13-19.

[17]. Karatekin, C., and Başaran, T. (2022) Gün *öncesi piyasasında elektrik enerjisi fiyatının veri analizi ile tahmin edilmesi*, Journal of the Institute of Science and Technology, 12(4): 2075-2084.

[18]. Web Page: Luke Hair Loss Dataset https://www.kaggle.com/datasets/lukexun/luke-hair-loss-dataset

[19]. Web Page: Hair loss dataset https://www.kaggle.com/datasets/brijlaldhankour/hair-loss-dataset

[20]. Krishnamoorthy, N., Jayanthi, P., Kumaravel, T., Sundareshwar, V. A., and Harris, R. S. J. (2023) *Scalp disease analysis using deep learning models,* Applied and Computational Engineering, 2, 1003-1009.

[21]. Ha, C., Go, T., and Choi, W. (2024) *Intelligent healthcare platform for diagnosis of scalp and hair disorders,* Applied Sciences, 14(5): 1734.

[22]. Wang, W. C., Chen, L. B., and Chang, W. J. (2018) *Development and experimental evaluation of machine-learning techniques for an intelligent hairy scalp detection system,* Applied Sciences, 8(6): 853.

[23]. Sokolova, M., and Lapalme, G. (2009) *A systematic analysis of performance measures for classification tasks,* Information Processing and Management, 45(4): 427–437.

[24]. Powers, D. M. W. (2011) *Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation,* Journal of Machine Learning Technologies, 2(1): 37–63.

[25]. Fawcett, T. (2006) *An introduction to ROC analysis,* Pattern Recognition Letters, 27(8): 861–874.

[26]. Chai, T., and Draxler, R. R. (2014) *Root mean square error (RMSE) or mean absolute error (MAE)?,* Geoscientific Model Development, 7(1): 1247–1250.

[27]. Yin, Y., and Jin, Y. (2020) *An improved R-squared statistic for evaluating regression models,* Journal of Applied Statistics, 47(6): 1114–1130.