International Journal of
AI EBIOMEDICINE INNOVATIONS
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0507 www.zesterapublications.com Original Research Paper

Hierarchical Probabilistic Learning for Tri-Class Hair Health Analytics and Hormonal Status Identification

Thatikayala Sreeja, E. Vishveshwar*
Vaagdevi Engineering College, Warangal, 506005, Telangana, India.
*Correspondence: E. Vishveshwar

ABSTRACT

Hair-related disorders have become increasingly prevalent due to factors such as unhealthy lifestyles, environmental pollution, genetic predisposition, nutritional deficiencies, and hormonal imbalances. Conventional diagnostic approaches primarily depend on clinical examination and expert interpretation, making early diagnosis challenging and limiting the ability to accurately evaluate the combined influence of multiple health-related factors. To overcome these limitations, this study proposes an Explainable Artificial Intelligence (XAI)-driven framework for intelligent hair health prediction using machine learning and deep learning techniques. The proposed system is implemented using the Flask web framework and incorporates comprehensive data preprocessing, exploratory data analysis, and comparative evaluation of several machine learning models, including Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbors (KNN), Gradient Boosting (GB), and AdaBoost (AB). In addition, a novel hybrid model combining a Probabilistic Neural Network (PNN) and a Sparse Representation Classifier (SRC) is developed to improve predictive performance. Experimental evaluation demonstrates that the proposed hybrid model achieves an accuracy of 0.9950, outperforming the baseline classifiers. To enhance model transparency and clinical interpretability, XAI techniques are integrated to identify the most influential prediction factors, estimate individual risk levels, and generate personalized treatment recommendations. Furthermore, the proposed framework supports multiple prediction tasks, including hair loss assessment, treatment recommendation, and hormonal impact analysis, providing an interpretable and reliable decision-support system for intelligent hair healthcare.

Keywords: Hair Health Prediction, Machine Learning, Probabilistic Neural Network (PNN), Sparse Representation Classifier (SRC), Explainable Artificial Intelligence (XAI)

1.INTRODUCTION

Scalp and hair disorders have become increasingly prevalent due to factors such as elevated stress levels, unhealthy dietary habits, hormonal imbalances, genetic predisposition, and continuous exposure to environmental pollutants. Common scalp conditions include dryness, excessive oil secretion, erythema, folliculitis, dandruff, and progressive hair loss, which, if left untreated, may develop into severe inflammatory disorders and permanent follicular damage [1]. These conditions not only affect physical appearance but also have a significant impact on an individual's confidence, emotional well-being, and social interactions. Consequently, timely diagnosis and appropriate treatment are essential for preventing disease progression and maintaining overall scalp and hair health [2]. The continuous growth of the global scalp and hair care market further reflects the increasing demand for intelligent, accurate, and technology-driven diagnostic solutions capable of supporting early intervention and personalized healthcare. Traditional diagnosis of scalp and hair disorders primarily depends on manual clinical examination performed by dermatologists or trained healthcare professionals. This process generally involves visual assessment of scalp conditions, patient history analysis, and subjective interpretation of symptoms, making the diagnosis highly

International Journal of
AI EBIOMEDICINE INNOVATIONS
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0507 www.zesterapublications.com Original Research Paper

dependent on the expertise and experience of the clinician. Such conventional approaches are often time-consuming, labor-intensive, and require patients to visit specialized healthcare facilities, limiting accessibility and scalability for large populations [3]. In addition, variations in clinical judgment may result in inconsistent diagnoses and treatment recommendations, while the dependence on highly trained specialists increases the overall cost and complexity of the diagnostic process.

Recent advances in AI and ML have created new opportunities for developing automated systems capable of improving the accuracy and efficiency of scalp and hair disorder diagnosis [4]. Several studies have explored computational approaches for detecting scalp abnormalities, identifying affected regions, assessing disease severity, and supporting clinical decision-making. Although these methods have demonstrated promising performance, many existing approaches focus on limited diagnostic tasks and are unable to simultaneously analyse multiple contributing factors, classify disorder severity, and provide comprehensive prediction within a unified framework [5]. Motivated by these limitations, the present study develops a predictive framework by evaluating multiple ML models, including SVM, LR, DT, KNN, GB, and AB, and introduces a hybrid PEDP model that combines PNN and SRC to improve predictive performance. The proposed framework is comprehensively evaluated using accuracy, precision, recall, and F1-score to demonstrate its effectiveness for intelligent hair disorder prediction and decision support.

2. LITERATURE SURVEY

Ahmad et al. [6] explored the application of ML techniques for the early prediction and assessment of hair loss, highlighting their potential in supporting personalized treatment strategies. The study evaluated multiple classification methods, including RF, SVM, and KNN, while incorporating feature engineering, data preprocessing, and hyperparameter optimization to improve predictive performance. Experimental findings demonstrated higher accuracy and precision compared to conventional approaches, emphasizing the effectiveness of automated diagnostic systems in enhancing hair loss detection and clinical decision-making.Ha et al. [7] introduced an intelligent healthcare platform for assessing the severity of six prevalent scalp disorders, including dryness, oiliness, erythema, folliculitis, dandruff, and hair loss. The study compared three deep learning architectures—ResNet-152, EfficientNet-B6, and ViT-B/16—to identify the most effective scalp image classification model. Among these, ViT-B/16 achieved the highest average classification accuracy of 78.31%. Furthermore, the attention rollout technique was employed to visualize lesion regions without requiring additional annotations. The developed scalp checker software enables objective scalp condition monitoring and supports the early diagnosis of scalp-related disorders.Kim et al. [8] proposed a multitask deep learning framework based on Mask R-CNN to classify hair follicles and estimate hair loss severity with high accuracy. Microscopic scalp images were resized, augmented, and processed using pretrained ResNet models for feature extraction. Hair follicle count, hair thickness, and the number of hairs per follicle were identified as the primary indicators for severity assessment. Using annotated images collected from individuals with different stages of hair loss, the framework performed instance segmentation and categorized follicle conditions into three predefined classes. Austin et al. [9] developed a non-invasive approach for identifying ASD biomarkers by integrating mass spectrometry analysis of elemental metabolism in hair strands with ML techniques. The investigation included prospective studies involving participants from Japan, Sweden, and the

United States. The predictive model successfully identified ASD risk as early as one month of age, achieving 96.4% sensitivity, 75.4% specificity, and 81.4% overall accuracy. The results demonstrated that abnormalities in elemental metabolism can be reliably detected through hair samples, offering a promising method for early ASD prediction.

Kim M. et al. [10] investigated the application of deep learning for automated HDM in hair-loss diagnosis. A dataset containing 4,492 enlarged RGB scalp images with detailed follicle annotations was utilized to train object detection models capable of identifying follicle locations and classifications. The study demonstrated that automated object detection can effectively support follicle counting and density estimation, improving the accuracy and efficiency of hair-loss assessment procedures. Hwang et al. [11] proposed a GAN-based ROI image translation framework to generate realistic post-hair transplantation predictions while preserving the non-ROI regions of the original image. The method independently performs image translation and segmentation, allowing only the target region to be modified. To further improve segmentation robustness, an ensemble strategy combining multiple segmentation models was incorporated. The proposed framework produced more reliable predictive visualizations and enhanced the accuracy of ROI localization.

Igarashi et al. [12] examined the relationship between GD and thyroid hormone concentrations in hair samples, demonstrating the feasibility of a non-invasive diagnostic approach. Using optimized LC-MS techniques, FT3 and FT4 were successfully detected from only 2 mg of hair with high sensitivity. Hair samples collected from GD patients exhibited significantly higher hormone concentrations than those of healthy individuals, closely corresponding with blood hormone levels. ROC analysis reported an AUC of 0.974 for FT3 and 0.900 for FT4, indicating excellent predictive performance and supporting the potential of hair-based hormone analysis for GD diagnosis. P. Duraisamy et al. [13] investigated the effectiveness of ML algorithms for predicting hair health by utilizing a comprehensive dataset that included individual characteristics and lifestyle factors. Multiple classification models, including LR, RF, and AB, were evaluated, with RF achieving the highest prediction accuracy of 94.6%. The study identified genetics, dietary habits, and stress as major contributors to hair health, highlighting the potential of intelligent predictive models for personalized and non-invasive hair care recommendations. Jin et al. [14] developed an intelligent web-based platform that enables users to assess scalp conditions through uploaded images. The diagnostic model achieved classification accuracies of 75% for fine dandruff and 82% for perifollicular erythema while effectively distinguishing normal, mild, moderate, and severe scalp conditions. The platform provides immediate diagnostic results, similar case references, and recommended solutions. User evaluations demonstrated high satisfaction regarding usability, interface design, and overall experience, emphasizing its practicality for accessible scalp health monitoring and early diagnosis. Nagase et al. [15] presented a comprehensive review of hair structure from both macroscopic and microscopic perspectives, explaining how structural characteristics influence the visual appearance of hair. The study described the role of optical phenomena, including reflection, refraction, scattering, and absorption, in determining hair appearance. Factors such as the alignment of hair fibers, cross-sectional geometry, and internal microstructures, including the cuticle, cortex, and medulla, were identified as key determinants of hair luster and texture. The review further explained that irregular fiber alignment, uplifted cuticles, and structural

International Journal of
AI EBIOMEDICINE INNOVATIONS
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0507 www.zesterapublications.com Original Research Paper

damage increase light scattering, resulting in reduced shine and a dull hair appearance.

3. PROPOSED SYSTEM

The proposed methodology establishes a systematic framework for predicting hair health by integrating data preprocessing, feature engineering, model training, and real-time prediction. Initially, the collected dataset undergoes preprocessing to handle missing values, eliminate inconsistencies, and encode categorical features into numerical representations suitable for machine learning. The processed data are then utilized to train and evaluate multiple conventional machine learning algorithms to assess their predictive capabilities. In addition, a hybrid deep learning and ensemble-based classifier is employed to enhance classification accuracy, robustness, and generalization performance, as illustrated in Figure .1. A lightweight database is incorporated to efficiently manage user profiles, prediction histories, and generated results. Furthermore, a web-based interface enables seamless user interaction by facilitating data input, visualization, and real-time hair health prediction. The proposed framework also supports periodic model retraining using newly acquired data, ensuring continuous performance improvement and adaptability to evolving data distributions.

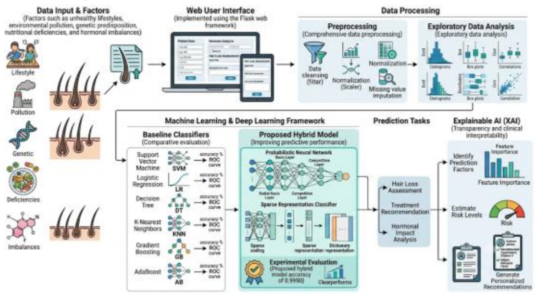


Figure 1: Proposed system architecture for hybrid probabilistic ensemble deep learning framework for hair disorder diagnosis.

Data Acquisition: Hair health-related information is collected from users through a Flask-based web application. The dataset comprises demographic information and multiple factors influencing hair health, including lifestyle habits, environmental pollution exposure, genetic predisposition, nutritional deficiencies, and hormonal imbalances. These attributes collectively form a comprehensive dataset for hair health assessment and prediction.

Data Preprocessing: The collected dataset undergoes preprocessing to improve data quality and consistency. Missing or inconsistent values are handled using suitable imputation techniques, while duplicate and irrelevant records are removed. Numerical features are normalized using a scaling technique, and categorical variables are transformed into numerical representations through Label Encoding or One-Hot Encoding to ensure compatibility with the predictive models.

Exploratory Data Analysis: EDA is performed to understand the statistical properties and relationships within the dataset. Histograms, distribution plots, box plots, and correlation analysis are employed to identify feature distributions, detect outliers, and evaluate inter-feature relationships. This analysis provides valuable insights into the characteristics of the data prior to model development.

Feature Extraction: Relevant features associated with hair health are selected from the pre-processed dataset. Both numerical and encoded categorical attributes are combined to construct a multidimensional feature space that effectively represents the factors influencing hair health. The extracted features serve as input for the predictive models.

PNN-SRC Hybrid Model: The extracted feature set is supplied to the proposed **PNN-SRC** hybrid framework. The PNN performs nonlinear probabilistic classification by estimating class membership probabilities, while the SRC represents input samples using sparse linear combinations of dictionary atoms to improve classification robustness. The integration of these two approaches enhances prediction accuracy, generalization capability, and resilience to noisy or incomplete data, resulting in reliable hair health prediction.

Baseline Model Evaluation: To assess the effectiveness of the proposed framework, several conventional ML classifiers, including SVM, LR, DT, KNN, GB, and AB, are trained and evaluated under identical experimental conditions. Comparative analysis is performed using standard evaluation metrics to demonstrate the superiority of the proposed hybrid model.

Feature Importance Analysis: An explainable AI module is incorporated to improve the transparency of the prediction process. Feature importance analysis identifies the relative contribution of individual attributes, such as lifestyle, genetics, nutritional deficiencies, hormonal imbalance, and environmental pollution, enabling users to understand the key factors influencing the predicted hair health condition.

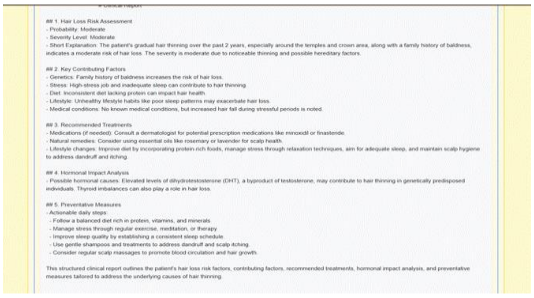
Model Deployment and Real-Time Prediction: The trained prediction model, preprocessing pipeline, and database are integrated into a Flask-based web application. User inputs are processed in real time to generate hair health predictions, assess hair loss severity, estimate associated risk levels, and provide personalized recommendations. The framework also stores prediction records and supports periodic model retraining with newly acquired data, ensuring continuous performance improvement and adaptability.

4. Results and Description

The research framework effectively integrates data acquisition, preprocessing, exploratory data analysis, predictive modeling, and explainable AI into a unified hair health prediction system. User information related to lifestyle, environmental pollution, genetic predisposition, nutritional deficiencies, and hormonal imbalance is collected through a Flask-based interface and processed using preprocessing techniques such as missing value handling, normalization, and categorical feature encoding to generate a high-quality dataset. The processed data are evaluated using baseline ML classifiers, including SVM, LR, DT, KNN, GB, and AB, followed by the proposed PNN-SRC hybrid model, which combines probabilistic learning with sparse representation to improve prediction accuracy, robustness, and generalization capability. The trained model is deployed within the web application to provide real-time hair health assessment, risk estimation, and personalized recommendations. Furthermore, the explainable AI module identifies the most influential factors contributing to each prediction, improving model transparency and supporting informed decision-making. Overall, the proposed framework provides an efficient, interpretable, and scalable solution for intelligent hair health prediction and personalized healthcare support.Figure 2 illustrates the Hair Health Prediction interface, where users provide personal and hair-related information to generate predictions. The collected inputs are processed by the predictive model to assess hair health and produce personalized, explainable results in real time.

Figure 2: Hair health XAI prediction

Figure 3 presents the XAI-based prediction report, displaying the predicted hair health status along with the corresponding risk level and explanation. It highlights the key factors influencing the prediction and provides personalized treatment recommendations and preventive measures to support informed hair care decisions.



Clinical Report

1. Hair Loss Risk Assessment
- Probability: Moderate
- Severity Level: Moderate
- Short Explanation: The patient's gradual hair thinning over the past 2 years, especially around the temples and crown area, along with a family history of baldness, indicates a moderate risk of hair loss. The severity is moderate due to noticeable thinning and possible hereditary factors.

2. Key Contributing Factors
- Genetics: Family history of baldness increases the risk of hair loss.
- Stress: High-stress job and inadequate sleep can contribute to hair thinning.
- Diet: Inconsistent diet lacking protein can impact hair health.
- Lifestyle: Unhealthy lifestyle habits like poor sleep patterns may exacerbate hair loss.
- Medical conditions: No known medical conditions, but increased hair fall during stressful periods is noted.

3. Recommended Treatments
- Medications (if needed): Consult a dermatologist for potential prescription medications like minoxidil or finasteride.
- Natural remedies: Consider using essential oils like rosemary or lavender for scalp health.
- Lifestyle changes: Improve diet by incorporating protein-rich foods, manage stress through relaxation techniques, aim for adequate sleep, and maintain scalp hygiene to address dandruff and itching.

4. Hormonal Impact Analysis
- Possible hormonal causes: Elevated levels of dihydrotestosterone (DHT), a byproduct of testosterone, may contribute to hair thinning in genetically predisposed individuals. Thyroid imbalances can also play a role in hair loss.

5. Preventative Measures
- Actionable daily steps:
  - Follow a balanced diet rich in protein, vitamins, and minerals.
  - Manage stress through regular exercise, meditation, or therapy.
  - Improve sleep quality by establishing a consistent sleep schedule.
  - Use gentle shampoos and treatments to address dandruff and scalp itching.
  - Consider regular scalp massages to promote blood circulation and hair growth.

This structured clinical report outlines the patient's hair loss risk factors, contributing factors, recommended treatments, hormonal impact analysis, and preventative measures tailored to address the underlying causes of hair thinning.

Figure 3: Detailed XAI prediction report

A comparative analysis was conducted to evaluate the performance of SVM, LR, DT, KNN, GB, AB, and the proposed PEDL model using accuracy, precision, recall, and F1-score. The comparison identifies the effectiveness of each classifier in predicting hair loss, medication & treatment recommendations, and hormonal changes.Table 1 presents the performance comparison for hair loss prediction. Among the conventional models, KNN achieved the highest accuracy of 0.7913, followed by GB (0.6725), while DT recorded the lowest accuracy (0.5112). The proposed PEDL model significantly outperformed all baseline models, achieving an accuracy, precision, recall, and F1-score of 0.9950, demonstrating its superior predictive capability.

Table 1: Comparative analysis of classification models for hair loss prediction

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.5463	0.5515	0.5463	0.5310
LR	0.5537	0.5546	0.5537	0.5529
DT	0.5112	0.5115	0.5112	0.5109
KNN	0.7913	0.7916	0.7913	0.7912
GB	0.6725	0.6726	0.6725	0.6725
AB	0.5700	0.5704	0.5700	0.5698
PEDL	0.9950	0.9950	0.9950	0.9950

Table 2 compares the models for medication and treatment prediction. KNN produced the best performance among the traditional classifiers with an accuracy of 0.6150, whereas SVM achieved the lowest accuracy (0.1200). The proposed PEDL model obtained the highest overall performance with an accuracy of 0.9862, consistently outperforming the existing methods. Table 3 summarizes the performance for hormonal changes prediction. KNN again achieved the highest accuracy (0.8100) among the conventional models, followed by GB (0.6737). The proposed PEDL model achieved the best overall results with an accuracy, precision, recall, and F1-score of 0.9938, confirming the effectiveness of the hybrid framework for accurate hair health prediction.

Table 2: Comparative analysis of classification models for medications & treatments prediction

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.1200	0.0486	0.1200	0.0587
LR	0.1613	0.1591	0.1613	0.1479
DT	0.1400	0.0999	0.1400	0.0654
KNN	0.6150	0.6161	0.6150	0.6137
GB	0.4800	0.4883	0.4800	0.4802
AB	0.1300	0.1372	0.1300	0.1266
PEDL	0.9862	0.9865	0.9862	0.9862

CONCLUSION

The proposed framework demonstrates the effectiveness of a data-driven approach for accurate hair health prediction by integrating data preprocessing, exploratory analysis, and predictive modeling. Comparative evaluation shows that KNN performs better than the other conventional classifiers, while the proposed PEDL model consistently achieves the highest performance across hair loss, medication and treatment, and hormonal changes prediction tasks. The hybrid integration of PNN and SRC enhances prediction accuracy, robustness, and generalization by effectively capturing complex feature relationships. Furthermore, the framework efficiently processes both categorical and numerical attributes, while evaluation metrics such as accuracy, precision, recall, and F1-score validate its reliability. The developed web-based system supports real-time prediction and personalized recommendations, making it a scalable and practical solution for intelligent hair health assessment and healthcare decision support.

REFERENCES

[1]. Kajiura, Y.; Watanabe, S.; Itou, T.; Nakamura, K.; Iida, A.; Inoue, K.; Yagi, N.; Shinohara, Y.; Amemiya, Y. Structural analysis of human hair single fibres by scanning microbeam SAXS. J. Struct. Biol. 2006, 155, 438–444.

[2]. Bryson, W.G.; Harland, D.P.; Caldwell, J.P.; Vernon, J.A.; Walls, R.J.; Woods, J.L.; Nagase, S.; Itou, T.; Koike, K. Cortical cell types and intermediate filament arrangements correlate with fiber curvature in Japanese human hair. J. Struct. Biol. 2009, 166, 46–58.

[3]. Harland, D.P.; Vernon, J.A.; Woods, J.L.; Nagase, S.; Itou, T.; Koike, K.; Scobie, D.A.; Grosvenor, A.J.; Dyer, J.M.; Clerens, S. Intrinsic curvature in wool fibres is determined by the relative length of orthocortical and paracortical cells. J. Exp. Biol. 2018, 221, 172312.

[4]. Bryson, W.G.; Harland, D.P.; Caldwell, J.P.; Vernon, J.A.; Walls, R.J.; Woods, J.L.; Nagase, S.; Itou, T.; Koike, K. Electron microscopy and tomography reveal that sodium 2-naphthalene sulfonate incorporated into perming solutions swells and tilts trichocyte intermediate filaments causing straightening of curly Japanese human hair. Int. J. Cosmet. Sci. 2019, 41, 132–146.

[5]. Nakamura, T.; Nagase, S.; Nojiri, M.; Breakspear, S.; Ishiwata, Y.; Fujita, A.; Hartmann, H. High definition dissociative delocalized dyes: A new class of physical hair dyes with unique 3D reflection on hair. In Proceedings of the 30th IFSCC Congress, Munich, Germany, 18–21 September 2018. O-S7-05.

[6]. Ahmad, M.; Mir, A.; Permana, A. An Early Hair Loss Detection and Prediction Method Based on Machine Learning. Eng. Proc. 2025, 107, 126. https://doi.org/10.3390/engproc2025107126

[7]. Ha, C.; Go, T.; Choi, W. Intelligent Healthcare Platform for Diagnosis of Scalp and Hair Disorders. Appl. Sci. 2024, 14, 1734. https://doi.org/10.3390/app14051734

[8]. Kim, J.-H.; Kwon, S.; Fu, J.; Park, J.-H. Hair Follicle Classification and Hair Loss Severity Estimation Using Mask R-CNN. J. Imaging 2022, 8, 283. https://doi.org/10.3390/jimaging8100283

[9]. Austin, C.; Curtin, P.; Arora, M.; Reichenberg, A.; Curtin, A.; Iwai-Shimada, M.; Wright, R.O.; Wright, R.J.; Remnelius, K.L.; Isaksson, J.; et al. Elemental Dynamics in Hair Accurately Predict Future Autism Spectrum Disorder Diagnosis: An International Multi-Center Study. J. Clin. Med. 2022, 11, 7154. https://doi.org/10.3390/jcm11237154

[10]. Kim, M.; Kang, S.; Lee, B.-D. Evaluation of Automated Measurement of

Hair Density Using Deep Neural Networks. Sensors 2022, 22, 650. https://doi.org/10.3390/s22020650

[11]. Hwang, D.-Y.; Choi, S.-H.; Shin, J.; Kim, M.; Choi, Y.-H. GAN-Based ROI Image Translation Method for Predicting Image after Hair Transplant Surgery. Electronics 2021, 10, 3066. https://doi.org/10.3390/electronics10243066

[12]. Igarashi, K.; Takita, C.; Matsumoto, M.; Kitagawa, W.; Ota, A.; Miyazaki, N.; Ito, K.; Ikeda, K. Application of Thyroid Hormones in Women's Hair for the Non-Invasive Prediction of Graves' Disease. Biomolecules 2025, 15, 353. https://doi.org/10.3390/biom15030353

[13]. P. Duraisamy, V. Vidhya, T. Preetha and V. Niranjani, "Implementing Machine Learning Techniques for Hair Health Prediction in Follicle Futures," 2024 Second International Conference on Data Science and Information System (ICDSIS), Hassan, India, 2024, pp. 1-6, doi: 10.1109/ICDSIS61070.2024.10594069

[14]. Jin, Y.-J.; Park, Y.-S.; Kang, S.-H.; Kim, D.-H.; Lee, J.-Y. A Study on the Development of a Web Platform for Scalp Diagnosis Using EfficientNet. Appl. Sci. 2024, 14, 7574. https://doi.org/10.3390/app14177574

[15]. Nagase, S. Hair Structures Affecting Hair Appearance. Cosmetics 2019, 6, 43. https://doi.org/10.3390/cosmetics6030043