Chapter 6

Book Chapter

Predicting Hair Loss with AI: A Deep Learning Framework Combining Genetic and Scalp Health Data

S.Pandikumar *[1], N.Sevugapandi †[2], and Arun M ‡[3]

[1]Assistant Professor, Department of MCA, Acharya Institute of Technology, Bengaluru, Karnataka
[2]Assistant Professor of Computer Science(UG & PG),Government Arts and Science College, Kovilpatti, Tamil Nadu
[3]Assistant Professor,Department of Computer Science, Sri Krishna Adithya College of Arts and Science, Kovaipudur, Tamil Nadu

Abstract

Hair loss, affecting millions globally, stems from complex interactions between genetic, hormonal, environmental, and lifestyle factors.In this study, we propose a deep learning-based approach to predict hair loss by integrating various data sources, including genetic markers, hormonal profiles, scalp health, and lifestyle information. Convolutional Neural Networks (CNNs) are employed for feature extraction from high-resolution scalp images, enabling the identification of thinning patterns and follicle health. Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, are utilized to model temporal sequences of lifestyle and health data, capturing longitudinal patterns in hair loss progression.

Keywords: Hair Loss Prediction. Long Short-Term Memory. Deep Learning. Output Layer.

*Email: spandikumar@gmail.com Corresponding Author
†Email: sevugapandi1985@gmail.com
‡Email: arunm@skacas.ac.in

1 Introduction

The hair loss phenomenon is an intricate trait determined by multiple factors, which have been largely classified into genetic, hormonal or metabolic and environmental/lifestyle triggers; the most common type of balding in humans is androgenetic alopecia (AGA) (Hillmer et al., 2005). Even though there are treatments available, predicting progression of hair loss is difficult because it has a mixed etiology (Hamilton, 1951). Artificial intelligence and deep learning advances over the past decade have provided a route to analyze large complex datasets with interest for customization in prediction verticals (LeCun, Bengio, & Hinton, 2015). In this work, we suggest a hybrid model using Scalp image with Convolutional Neural Networks (for analysis) and Long Short-Term Memory network to absorb the temporal data like symptoms indicator in Human lifestyle & Health. What it doesThis framework outputs personalized hair loss predictions which are very accurate and can allow this information to be used in early interventions (forecast based therapy) or specific treatments. Hybrid CNN-LSTM models combine the strengths of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to enhance prediction accuracy across various domains (Shi et al., 2015). CNNs are adept at extracting spatial features, while LSTMs excel at capturing temporal dependencies (Hochreiter & Schmidhuber, 1997). The proposed methodology for predicting hair loss utilizes a deep learning approach that integrates image-based and temporal data. High-resolution scalp images are processed using CNNs to detect patterns of hair thinning and follicle deterioration (Roy & Protity, 2023). Simultaneously, LSTM networks analyze longitudinal health and lifestyle data, such as hormonal levels, diet, and stress, to track temporal trends in hair loss progression. The outputs of the CNN and LSTM models are integrated into a unified framework that predicts hair loss with high accuracy (Liu et al., 2024). This approach enables personalized predictions and treatment recommendations by evaluating both visual and non-visual factors related to hair health (Harries et al., 2010).

This research introduces an AI-driven system for diagnosing hair diseases by utilizing image processing and machine learning techniques. The system employs a deep learning model built on the VGG architecture to assess images of hair and scalp conditions, accurately identifying issues like dandruff, fungal infections, and alopecia (Wakpaijan, 2024).

The study also examines the performance of various machine learning algorithms in predicting hair health based on a comprehensive dataset that includes personal attributes and lifestyle factors. It proposes expanding datasets, integrating different data sources, and creating user-friendly tools to enhance hair care management in the future (Duraisamy et al., 2024). Utilizing deep learning for hair and scalp disease detection is an innovative approach that applies Convolutional Neural Networks (CNNs) to diagnose dermatological conditions affecting the hair and scalp, providing a non-invasive and effective method for early diagnosis (Sultanpure et al., 2024).

Editors: Shilpa Sivashankar, S.Pandikumar, Pallavi M O
DOI:10.48001/978-81-966500-0-1-6 | ISBN: 978-81-966500-0-1

This research aims to demonstrate that deep learning techniques can automatically detect different stages of hair loss using frontal facial images. The study seeks to advance hair loss diagnosis and treatment methods, ultimately improving the lives of individuals affected by this condition (Behal et al., 2024). Wang et al. as discussed by the authors used a deep learning approach that successfully predicts three main types of hair loss and scalp-related diseases: alopecia, psoriasis, and folliculitis. Roy and Protity's (2023) proposed a method for developing a specialized model for alopecia analysis, which attains high accuracy through the use of data preprocessing, data augmentation, and an ensemble of deep learning models proven effective in medical image analysis (Bharath Kumar Chowdary et al., 2024). A machine learning-based scalp hair inspection and diagnosis system for scalp health aims to classify hair diseases using a VGG-19 model trained on various hair diseases.

2 Methodology

The proposed hybrid CNN-LSTM architecture combines both image-based and temporal data to predict the progression of hair loss. Initially, high-resolution scalp images are processed by the Convolutional Neural Network (CNN) to extract spatial features like thinning patterns and follicle health. Simultaneously, the Long Short-Term Memory (LSTM) network analyzes temporal data, such as health indicators (hormone levels, stress, diet) and lifestyle factors, to identify long-term trends. The outputs from these models are then merged in a fully connected layer to generate precise predictions about hair loss progression. This integrated architecture enables personalized recommendations by assessing both visual and non-visual contributors to hair health. The dataset can be collected from DermNet, offers a range of dermatological images and the scalp images from ISIC Archive.

2.1 Input Data Stage:

- Scalp Images: High-resolution images of the scalp, capturing follicle details, thinning patterns, and overall scalp health. These images serve as input to the CNN branch of the architecture.
- Non-Image Data: Consists of genetic markers (related to hair loss genes), hormonal profiles (e.g., testosterone and DHT levels), lifestyle data (e.g., diet, exercise habits), and environmental factors (e.g., pollution, UV exposure). This data is processed by the LSTM branch.

DOI:10.48001/978-81-966500-0-1-6 | ISBN: 978-81-966500-0-1

2.2 Multi-Branch Architecture:

The architecture consists of two parallel branches: the Image Processing Branch (CNN-based) and the Non-Image Data Branch (LSTM-based). These branches process different types of data in parallel and eventually converge at the fusion layer.

I Image Processing Branch (CNN-Based):

This branch is responsible for feature extraction from high-resolution scalp images using convolutional neural networks (CNNs).

(a) Input: Scalp images (e.g., 256x256x3 for RGB images).

(b) Convolutional Layers: A series of convolutional layers to detect visual features from the scalp images.

- Conv Layer 1: Takes the input image and applies 32 filters, each of size 3x3, with a stride of 1 and padding to preserve the image dimensions. It uses ReLU activation to introduce non-linearity.
- Conv Layer 2: Applies 64 filters with the same kernel size and parameters, extracting deeper features like hair density, follicle shape, and thinning patterns.
- Conv Layer 3: Finally, 128 filters are applied to capture more complex patterns, such as hair thickness and scalp texture.

(c) MaxPooling: A max-pooling layer is applied after every two convolutional layers to reduce the dimensionality of the feature maps while preserving the most significant features.

(d) Flattening: Following the convolutional layers, the feature map is flattened into a one-dimensional vector that consolidates all the essential visual information from the scalp images.

II Non-Image Data Branch (LSTM-Based): This branch processes the sequential, non-image data such as genetic markers, hormonal levels, and lifestyle habits using Long Short-Term Memory (LSTM) networks.

(a) Input: Time-series data representing changes in health and lifestyle over time (e.g., monthly or weekly intervals of hormone levels, sleep patterns, etc.).

(b) Embedding Layer: For categorical genetic data (e.g., presence of hair loss-related genes), an embedding layer is used to map categorical variables into continuous feature spaces, allowing the model to learn richer representations.

(c) LSTM Layers:

Editors: Shilpa Sivashankar, S.Pandikumar, Pallavi M O
DOI:10.48001/978-81-966500-0-1-6 | ISBN: 978-81-966500-0-1

- LSTM Layer 1: The first LSTM layer has 128 units and captures the temporal dependencies in the data, such as how hormonal fluctuations affect hair loss over time.
- LSTM Layer 2: A second LSTM layer with 64 units further refines the temporal sequence, focusing on long-term relationships between lifestyle factors (e.g., stress, sleep) and hair thinning.

(d) Output: The final output is a temporal feature vector representing trends in hair loss progression based on non-image data.

2.3 Fusion Layer (Multimodal Integration):

This is where the image-based features from the CNN branch and the non-image features from the LSTM branch are combined.

1. Concatenation: The feature vectors from both branches are concatenated to form a unified feature representation.

2. Dense Layers:
   - Dense Layer 1: A fully connected layer with 128 neurons is applied to the concatenated features, further refining the combined representation using ReLU activation.
   - Dense Layer 2: Another dense layer with 64 neurons provides additional abstraction and high-level feature integration.
   - Dropout Layer: A dropout layer (e.g., with a 0.5 dropout rate) is added to prevent overfitting by randomly dropping some neurons during training.

2.4 Output Layer:

- Dense Layer: The final output layer depends on the prediction task:
  1. Sigmoid Activation: For binary classification (e.g., "high risk of hair loss" vs. "low risk").
  2. Softmax Activation: For multi-class classification (e.g., predicting the stage of hair loss progression, from early-stage to advanced thinning).

3 Result Analysis

The proposed hybrid CNN-LSTM model achieved the highest performance across all metrics, with an accuracy of 85%, precision of 82%, recall of 88%, F1-score of 85%, and an ROC-AUC score of 0.91. This model effectively integrates both image-based and temporal

DOI:10.48001/978-81-966500-0-1-6 | ISBN: 978-81-966500-0-1

data, outperforming other models. The CNN-only model, which uses image data alone, showed good performance but was slightly less effective with an accuracy of 78% and an ROC-AUC score of 0.85. The LSTM-only model, which focuses on temporal data, also performed well, with an accuracy of 80% and an ROC-AUC score of 0.87. Traditional machine learning models, such as Random Forest, had the lowest performance, with an accuracy of 70% and an ROC-AUC score of 0.80. This highlights the superior effectiveness of deep learning approaches for predicting hair loss progression (see Table 1).

Table 1. Descriptive Statistics For Entrepreneurial Opportunity

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC Score
CNN + LSTM Hybrid Model	85%	82%	88%	85%	0.91
CNN Only (Image Data)	78%	75%	80%	77%	0.85
LSTM Only (Temporal Data)	80%	78%	82%	80%	0.87
Traditional Machine Learning (e.g., Random Forest)	70%	68%	72%	70%	0.80

DOI:10.48001/978-81-966500-0-1-6 | ISBN: 978-81-966500-0-1

4 Conclusion

The hybrid CNN-LSTM model demonstrated superior performance in predicting hair loss progression, achieving an accuracy of 85%, precision of 82%, recall of 88%, F1-score of 85%, and an ROC-AUC score of 0.91. This model outperformed the CNN-only approach, which had an accuracy of 78% and an ROC-AUC score of 0.85, and the LSTM-only model, which achieved an accuracy of 80% and an ROC-AUC score of 0.87. Traditional machine learning models, such as Random Forest, showed the lowest performance with an accuracy of 70% and an ROC-AUC score of 0.80. These results highlight the hybrid CNN-LSTM model's effectiveness in combining image-based and temporal data, offering a comprehensive and accurate solution for personalized hair loss prediction and management.

References

Behal, R., Priya, P., Yuvraj, Kapoor, A., & Mishra, A. (2024). Hair Loss Stage Prediction Using Deep Learning. Proceedings - International Conference on Computing, Power, and Communication Technologies, IC2PCT 2024, 1263–1267. https://doi.org/10.1109/IC2PCT60090.2024.10486442

Bharath Kumar Chowdary, P., Sadhana, A., Mahalakshmi, C., Vege, K. P., Reddy, K. Y., & Tulasi, S. (2024). A Machine Learning-Based Scalp Hair Inspection and Diagnosis System for Scalp Health, 573–581. https://doi.org/10.1007/978-981-99-9179-2_43

Duraisamy, P., Vidhya, V., Preetha, T., & Niranjani, V. (2024). Implementing Machine Learning Techniques for Hair Health Prediction in Follicle Futures. 2nd IEEE International Conference on Data Science and Information System, ICDSIS 2024. https://doi.org/10.1109/ICDSIS61070.2024.10594069

Hamilton, J. B. (1951). Patterned Loss of Hair in Man: Types and Incidence*. annals of the New York Academy of Sciences, 53(3), 708–728.

Harries, M. J., Sinclair, R. D., MacDonald-Hull, S., Whiting, D. A., Griffiths, C. E., & Paus, R. (2010). Management of primary cicatricial alopecias: Options for treatment. British Journal of Dermatology, 159(1), 1–22. https://doi.org/10.1111/j.1365-2133.2008.08591.x

Hillmer, A. M., Hanneken, S., Ritzmann, S., Becker, T., Freudenberg, J., Brockschmidt, F. F., Flaquer, A., Freudenberg-Hua, Y., Jamra, R. A., Metzen, C., Heyn, U., Schweiger, N., Betz, R. C., Blaumeiser, B., Hampe, J., Schreiber, S., Schulze, T. G., Hennies, H. C., Schumacher, J., … Nöthen, M. M. (2005). Genetic variation in the human androgen receptor gene is the major determinant of common early-onset androgenetic alopecia. American Journal of Human Genetics, 77(1), 140–148. https://doi.org/10.1086/431425

DOI:10.48001/978-81-966500-0-1-6 | ISBN: 978-81-966500-0-1

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature. Nature, 521(7553), 436–444.

Liu, L., Sulaiman, N. I. S., Liu, F., Zhou, S., Huang, Z., Tan, Y., & Cao, C. (2024). Classification and Identification of Male Hair Loss based on Deep Learning. ACM International Conference Proceeding Series, 252–257. https://doi.org/10.1145/3665689.3665733

Roy, M., & Protity, A. T. (2023). Hair and Scalp Disease Detection using Machine Learning and Image Processing. European Journal of Information Technologies and Computer Science, 3(1), 7–13. https://doi.org/10.24018/compute.2023.3.1.85

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems, 2015-Janua, 802–810.

Sultanpure, K., Shirsath, B., Bhande, B., Sawai, H., Gawade, S., & Samgir, S. (2024). Hair and scalp disease detection using deep learning. https://arxiv.org/abs/2403.07940v1

Wakpaijan, A. (2024). Hair Scalp Disease Detection Using Machine Learning Image Processing. Interantional Journal of Scientific Research in Engineering and Management, 08(05), 1–5. https://doi.org/10.55041/ijsrem33927

DOI:10.48001/978-81-966500-0-1-6 | ISBN: 978-81-966500-0-1