Volume 4, Issue 3, 2026

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

MACHINE LEARNING–BASED PREDICTION OF GRAFT SURVIVAL AND HAIR DENSITY AFTER HAIR TRANSPLANTATION USING MULTIMODAL CLINICAL AND TRICHOSCOPIC DATA

Dr. Muhammad Usman Amiruddin*[1], Shoaib Khan[2], Muhammad Harris Ijaz[3]

*[1]*Consultant Plastic Surgeon, MBBs, FCPS (Plastic), UA Aesthetics Lahore, Pakistan*
[2,3]*SEGI University, Malaysia*

*[1]usman_adin@hotmail.com, [2]shoaib5192071@gmail.com, [3]harrisrana08@gmail.com

DOI: https://doi.org/10.5281/zenodo.23317346

Keywords

Machine learning; hair transplantation; graft survival prediction; hair density estimation; multimodal clinical data; trichoscopic image analysis; ridge regression; random forest

Article History

Received: 19 January 2026
Accepted: 04 March 2026
Published: 18 March 2026

Copyright @Author

Corresponding Author: *
Dr. Muhammad Usman Amiruddin

Abstract

Background: *Accurate prediction of follicular graft survival and postoperative hair density remains a major challenge in hair transplantation because of interpatient variability, donor follicle characteristics, recipient-site conditions, surgical parameters, and postoperative biological responses. Conventional predictive approaches may inadequately capture nonlinear relationships between clinical variables and trichoscopic features, limiting individualized outcome estimation. This study investigates a machine learning–based framework designed to integrate multimodal clinical and trichoscopic information for predicting hair-transplantation outcomes.*

Objective: *The primary objective is to develop and evaluate predictive models for graft survival and postoperative hair density and to investigate the potential contribution of integrating clinical characteristics, surgical variables, and trichoscopic features. The study also evaluates model performance using complementary regression and classification metrics to establish a quantitative foundation for personalized outcome assessment.*

Methods: *The analytical framework included 500 patient records and 23 variables, covering clinical characteristics, transplantation-related parameters, and postoperative outcome measures. Ridge regression and random forest algorithms were evaluated alongside clinical-only and trichoscopy-only predictive approaches. Model performance was assessed using mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R2R^2), accuracy, sensitivity, specificity, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). For graft-survival classification, a threshold of 85% graft survival was used to distinguish the outcome categories. Regression models were evaluated for continuous graft-survival and hair-density prediction, while classification metrics characterized the ability to identify patients meeting the specified graft-survival threshold.*

Results: *The ridge regression baseline achieved an MAE of 5.61 percentage points and an RMSE of 7.66 percentage points for graft-survival prediction, with an R2R^2 of 0.295. For postoperative hair-density prediction, the model achieved an MAE of 6.55 hairs/cm², an RMSE of 8.60 hairs/cm², and an R2R^2 of 0.758. The graft-survival classification model achieved a ROC-AUC of 0.776, an accuracy of 72.0%, a sensitivity of 86.2%, a specificity of 52.4%, an F1-score of 0.781, and a Brier score of 0.184. Comparative evaluation showed that the multimodal random forest achieved an MAE of 6.09 percentage points for graft survival and 8.10 hairs/cm² for hair density,*

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

whereas the clinical-only random forest achieved corresponding MAEs of 6.14 percentage points and 7.77 hairs/cm². These comparisons indicate that predictive performance differed by outcome and model configuration, emphasizing the importance of rigorous comparative evaluation rather than assuming that multimodal integration necessarily improves every endpoint.

Conclusion: *The findings demonstrate the potential of machine learning to quantify postoperative hair-transplantation outcomes and provide a foundation for individualized graft-survival and hair-density prediction. The observed R2R^2 of 0.758 for hair-density prediction indicates stronger model fit than the corresponding graft-survival R2R^2 of 0.295, highlighting differences in the predictability of the two outcomes within the evaluated analysis. The classification ROC-AUC of 0.776 further suggests useful discrimination potential, although the specificity of 52.4% indicates that classification performance requires improvement before clinical application. Further validation using real-world patient cohorts, independent external datasets, standardized postoperative measurements, and prospective multicentre studies is necessary to establish clinical reliability and generalizability. This research provides a quantitative starting point for developing interpretable, patient-centred decision-support systems in precision hair restoration medicine.*

1. Introduction

Hair loss, particularly androgenetic alopecia (AGA), can affect appearance, self-image, and quality of life. The clinical phenotype is shaped by genetic susceptibility, age, androgen-related follicular miniaturization, and variation in the distribution and caliber of terminal hairs. Early descriptions of the natural history of patterned hair loss by Hamilton (1951), and later classification systems by Norwood (1975) and Ludwig (1977), provided a common clinical language for describing the distribution and severity of hair loss. These systems remain useful for clinical documentation, but categorical grades alone do not fully represent the complex variation in donor capacity, recipient-zone characteristics, hair caliber, follicular-unit composition, and postoperative growth that can influence an individual patient's result.

Hair transplantation has evolved from early grafting procedures to follicular-unit-based methods designed to preserve natural follicular groupings and create a more natural appearance. Orentreich (1959) established the concept of donor dominance, a foundational principle in surgical hair restoration. Bernstein and Rassman (1999) described the rationale for follicular-unit transplantation, emphasizing preservation of naturally occurring follicular units. Subsequent discussions by Unger (2000), Bernstein, Rassman, and Stough (2000), Rassman and colleagues (2002), and Bernstein and Rassman (2005, 2006) addressed surgical technique, follicular-unit extraction, graft handling, and postoperative graft anchoring. These contributions established important procedural principles, while also highlighting that outcomes depend on several interacting biological, technical, and patient-level factors.

Despite advances in surgical technique, preoperative counseling still involves uncertainty about the degree of graft survival and the hair density that will be visible after transplantation. Patients with similar clinical grades or graft counts may have different results because of differences in donor density, shaft diameter, miniaturization, scalp characteristics, recipient-area size, surgical handling, concurrent medical treatment, and follow-up duration. A simple average or a single clinical score may not adequately capture these interactions. A quantitative prediction approach could help organize relevant evidence, support individualized counseling, and identify cases requiring closer follow-up. Such a model should support—not replace—clinical judgment, and should not be used for treatment decisions until it has been adequately evaluated in representative clinical data.

Trichoscopy, using dermoscopy or videodermoscopy to examine hair and scalp

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

structures, offers a non-invasive method for quantifying features that may be difficult to capture through visual grading alone. Rakowska and colleagues (2009) described standardized dermoscopic criteria for female pattern hair loss, while Rakowska (2009) discussed standardization and measurable parameters in healthy scalp imaging. Tosti and Duque-Estrada (2009) reviewed dermoscopy in hair disorders, and Rudnicka and colleagues (2011) summarized the expanding diagnostic role of trichoscopy. Inui (2011) further discussed trichoscopy in the assessment of alopecia. These studies support the value of structured scalp imaging, but diagnostic characterization of alopecia is not the same task as forecasting graft survival or postoperative hair density after surgery.

Clinical research on AGA has also developed through treatment reviews and evidence-based guidance. Kaufman and colleagues (1998) evaluated finasteride in men with AGA, while Olsen and colleagues (2002) reported clinical-trial evidence for topical minoxidil in male-pattern hair loss. Adil and Godwin (2017) reviewed evidence for medical treatments, Lolli and colleagues (2017) summarized the pathogenesis and treatment of AGA, and Suchonwanit and colleagues (2019) provided an updated review of AGA management. Kanti and colleagues (2018) developed evidence-based European guidance for AGA. These works demonstrate that postoperative appearance exists within a broader treatment context: concurrent therapies, baseline disease activity, and patient characteristics should therefore be recorded when developing a prediction model.

Machine learning provides methods for representing nonlinear relationships and interactions across mixed clinical and image-derived predictors. Breiman (2001) introduced random forests as an ensemble-learning approach, and Friedman (2001) developed gradient boosting methods for predictive modeling. These algorithms can complement simpler regression models, but their flexibility also increases the risk of overfitting when the number of independent patients is limited. Interpretation methods such as LIME, described by Ribeiro, Singh, and Guestrin (2016), and SHAP, developed by Lundberg and Lee (2017), can help explain model outputs, although explanations do not establish causality. For clinical prediction studies, Steyerberg (2019) emphasized careful model development and validation; Collins and colleagues (2015) introduced TRIPOD reporting guidance; Wolff and colleagues (2019) presented PROBAST for assessing risk of bias and applicability; Riley and colleagues (2020) described sample-size considerations for prediction models; and Van Calster and colleagues (2019) highlighted the central importance of calibration. Most recently, Collins and colleagues (2024) published TRIPOD+AI guidance for transparent reporting of prediction models developed using regression or machine-learning methods.

Against this background, the present study proposes a multimodal framework that combines structured clinical and surgical information with quantitative trichoscopic features to predict two clinically relevant continuous outcomes: graft survival percentage and postoperative hair density in hairs/cm^2. The study will compare clinical-only, trichoscopy-only, and combined models, use patient-level validation to prevent data leakage, and evaluate both predictive error and calibration. The intended contribution is a carefully evaluated prediction framework that links standardized clinical assessment with objective image-derived measurements, while explicitly distinguishing exploratory model development from evidence of clinical utility.

Hamilton (1951), Norwood (1975), and Ludwig (1977) provided foundational descriptions of patterned hair loss and its clinical distribution. These classifications make cohorts easier to describe and compare, but they were not designed as individualized prognostic algorithms for surgical outcomes. A patient's visible pattern is only one dimension of the transplant-planning problem; it does not independently encode donor supply, the caliber and grouping of hairs, recipient-zone surface area, or procedure-specific variables. Consequently, a modern prediction study should retain standardized clinical grades while also recording continuous measurements and relevant treatment history.

Orentreich (1959) established donor dominance as a core biological principle. Bernstein and Rassman (1999) described the

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

logic of transplanting follicular units, and Unger (2000) debated the scope and limits of follicular-unit techniques. Bernstein, Rassman, and Stough (2000) supported the follicular-unit approach, while Rassman and colleagues (2002) described follicular-unit extraction as a minimally invasive method. Bernstein and Rassman (2005) reviewed the state of follicular-unit transplantation, and their later work on graft anchoring (2006) examined early postoperative graft stability. Avram and Rogers (2009) reviewed contemporary hair-transplantation practice and complications. Collectively, this literature establishes the technical and biological foundations of transplantation, but clinical procedure descriptions do not by themselves provide a validated multivariable prediction tool for individual graft survival and final density.

Rakowska and colleagues (2009) proposed standardized dermoscopic criteria for female androgenetic alopecia, and Rakowska (2009) reported methods for standardizing trichoscopic measurements and establishing reference values. Tosti and Duque-Estrada (2009) described dermoscopy as a useful adjunct for hair-disorder assessment. Rudnicka and colleagues (2011) reviewed trichoscopic applications across non-scarring and scarring alopecia, hair-shaft disorders, and inflammatory scalp disease; Inui (2011) also summarized its clinical diagnostic role. The common methodological implication is that image acquisition and interpretation require standardization. Magnification, scalp location, lighting, image quality, and the method used to count or estimate hairs can influence the measured features. For prediction research, these factors should be treated as protocol variables and quality-control targets, not as incidental details.

Kaufman and colleagues (1998) reported trial evidence for finasteride in men with AGA, and Olsen and colleagues (2002) evaluated topical minoxidil. Adil and Godwin (2017) systematically reviewed medical treatments for AGA, Lolli and colleagues (2017) summarized disease mechanisms and treatment options, Suchonwanit and colleagues (2019) reviewed current management, and Kanti and colleagues (2018) produced evidence-based European guidance. These studies are relevant because postoperative hair density may be influenced not only by transplanted follicles but also by native hair miniaturization, the timing of concurrent therapy, and the underlying progression of hair loss. A predictive dataset should record these factors and specify which information is available at the intended prediction time. Postoperative variables must not be used to predict outcomes that are intended to be forecast preoperatively.

Breiman (2001) introduced random forests, while Friedman (2001) developed gradient boosting for predictive learning. These approaches can model nonlinear associations and interactions among clinical and image-derived variables. However, performance estimates can be optimistic if preprocessing, feature selection, or hyperparameter tuning uses information from the test set. Steyerberg (2019) provides a comprehensive framework for developing and validating clinical prediction models. Collins and colleagues (2015) proposed TRIPOD reporting recommendations, and Wolff and colleagues (2019) developed PROBAST to assess risk of bias and applicability. Riley and colleagues (2020) emphasized that sample-size planning must reflect the intended model, outcome distribution, and desired precision. Van Calster and colleagues (2019) stressed that calibration—the agreement between predicted and observed outcomes—is essential alongside discrimination or error metrics. Ribeiro and colleagues (2016) and Lundberg and Lee (2017) introduced explanation approaches that may aid model interpretation, but such explanations should be presented as model-behavior summaries rather than causal conclusions. Collins and colleagues (2024) updated reporting guidance through TRIPOD+AI, explicitly covering regression and machine-learning prediction studies.

The literature therefore provides strong foundations in surgical technique, scalp phenotyping, hair-loss treatment, and general predictive modeling. These strands do not automatically combine into a reliable postoperative outcome model. A study focused on graft survival and postoperative hair density must establish a reproducible outcome definition, link baseline clinical and trichoscopic predictors to a clearly specified endpoint window, prevent patient-level data

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

leakage, compare unimodal and multimodal approaches, quantify uncertainty, and evaluate calibration before considering clinical implementation.

The central research gap is the limited integration of standardized clinical/surgical variables and quantitative trichoscopic measurements into a single, rigorously evaluated model specifically designed to predict graft survival percentage and postoperative hair density after hair transplantation. The literature reviewed here establishes the importance of follicular-unit biology and surgical technique (Orentreich, 1959; Bernstein and Rassman, 1999; Rassman et al., 2002), the utility of trichoscopy for scalp phenotyping (Rakowska et al., 2009; Rudnicka et al., 2011), and the methodological requirements for clinical prediction models (Steyerberg, 2019; Riley et al., 2020; Collins et al., 2024). However, these contributions address related but distinct problems. A diagnosis-oriented trichoscopy study does not directly answer how well an individual will respond to transplantation, and a technical review of transplantation does not establish an individualized, validated prediction model.

A second gap concerns outcome measurement. "Graft survival" and "postoperative hair density" are related but not interchangeable. A study must define whether graft survival can be counted directly, specify the numerator and denominator, and distinguish survival of transplanted grafts from visible hair density that may also reflect native hairs. Hair density requires calibrated image area, reproducible scalp zones, consistent magnification, and a prespecified follow-up window. Without clear endpoint definitions and reliable measurement, even a sophisticated algorithm may learn measurement artifacts rather than clinically meaningful outcomes.

A third gap concerns multimodal incremental value. Clinical variables may capture patient context and surgical planning, while trichoscopy may quantify hair density, shaft caliber, miniaturization, and follicular-unit composition. It remains an empirical question whether combining these modalities improves prediction beyond clinical data alone or trichoscopy alone. The proposed study addresses this by comparing clinical-only, trichoscopy-only, and multimodal models against a transparent baseline, rather than assuming that more inputs necessarily improve performance.

A fourth gap is methodological robustness and transportability. Small cohorts, repeated images from the same person, inconsistent acquisition settings, incomplete follow-up, and leakage during feature selection or preprocessing can produce overoptimistic results. The proposed framework will use patient-level partitions, training-only preprocessing, nested tuning or cross-validation when appropriate, uncertainty intervals, residual and calibration analyses, and—if feasible—external or temporal validation. Reporting will be aligned with current TRIPOD+AI recommendations, and risk-of-bias considerations will be made explicit.

Finally, a clinically useful model must be interpretable and responsibly deployed. Feature-importance measures can help describe which variables influence predictions, but they do not prove causal effects. Subgroup performance, missing-data patterns, device differences, and center effects should be examined where sample size permits. External validation and prospective impact evaluation would be required before using model predictions to guide care. Accordingly, this research aims to contribute not merely a high-performing algorithm, but a transparent multimodal prediction framework with clinically interpretable endpoints and a validation plan suited to real-world hair-transplant practice.

Overall aim: To develop and internally evaluate machine-learning models that predict graft survival percentage and postoperative hair density after hair transplantation using multimodal clinical, surgical, and trichoscopic data.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Table 1: Alignment of research gap, objective, and evaluation

Research gap	Study objective	Planned evaluation
Limited integration of clinical and trichoscopic data	Develop clinical-only, trichoscopy-only, and multimodal models	Compare MAE, RMSE, R^2, and uncertainty intervals
Potential ambiguity in outcomes	Define graft survival and hair density using prespecified measurement protocols	Quality control, endpoint audit, assessor blinding where feasible
Risk of overfitting and leakage	Use patient-level splits and training-only preprocessing	Nested cross-validation or locked test set
Limited interpretability	Describe influential predictors without causal overclaiming	Feature importance / SHAP-style summaries plus clinical review
Uncertain generalizability	Assess transportability where data permit	Temporal, site-based, or external validation

Research questions: (1) How accurately can graft survival percentage be predicted from baseline clinical and trichoscopic data? (2) How accurately can postoperative hair density be predicted? (3) Does combining clinical and trichoscopic predictors improve performance compared with either modality alone? (4) Which features contribute most consistently to model predictions? (5) How well calibrated and transportable are the predictions across patient subgroups, clinics, or time periods, where sample size permits? The expected contribution is a structured evaluation of multimodal prediction for two distinct postoperative outcomes, with explicit attention to outcome validity, imaging standardization, leakage prevention, model comparison, and transparent reporting. If supported by real clinical data and external evaluation, such a framework could assist individualized counseling, support audit of outcome variation, and help identify patients who may benefit from closer follow-up. Until those evaluations are completed, the model should be regarded as a research tool rather than a clinically validated decision aid.

Methodology

3.1 Research design and methodological approach

A prospective observational cohort design is proposed to develop and internally validate machine-learning models for predicting two continuous outcomes after hair transplantation: (i) graft survival percentage and (ii) postoperative hair density, expressed as hairs/cm^2. The study will integrate structured clinical variables with quantitative features derived from standardized trichoscopic images. A prospective design is preferred because predictor definitions, imaging conditions, endpoint timing, and follow-up can be standardized before outcomes are observed. If only retrospective records are available, the design should be explicitly relabeled retrospective and the limitations of missingness, selection bias, and inconsistent image acquisition should be reported. The unit of analysis will be the patient, with scalp zone and visit-level measurements retained as repeated observations where appropriate. Data from the same patient must never be distributed across training and test partitions. The primary aim is prediction rather than causal inference; therefore, model coefficients and feature importance will not be interpreted as evidence that a predictor causes improved graft survival.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

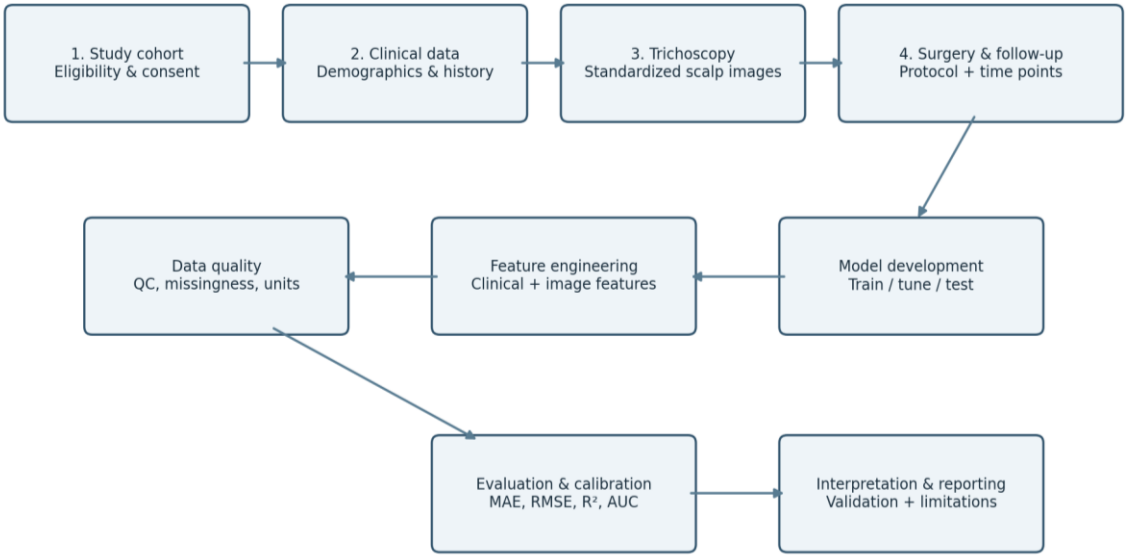


Figure 3.1: Proposed end-to-end workflow from cohort recruitment to model evaluation

3.2 Study setting, population, and eligibility

The study should be conducted at one or more hair-transplant clinics using a harmonized protocol. The target population is adults undergoing a clinically indicated hair-transplant procedure, subject to local ethics approval and the treating clinician's eligibility assessment. Recruitment should be consecutive where feasible to reduce discretionary selection. Before data collection, the protocol must define the transplant technique, treatment setting, follow-up window, and whether revision procedures or combined medical treatments are included.

Table 3.1: Proposed eligibility and exclusion criteria

Domain	Proposed operational rule	Reason / implementation note
Inclusion	Adult patients undergoing the specified hair-transplant procedure; written informed consent; baseline clinical assessment and usable baseline trichoscopy	Confirm age threshold and procedure details in the approved protocol
Follow-up	At least one prespecified primary endpoint visit in the chosen 9–12-month window	Record missed visits and reasons; do not silently treat them as negative outcomes
Image quality	Images meet predefined focus, magnification, lighting, and scalp-zone requirements	Repeat acquisition when possible; document non-recoverable failures
Exclusion	No valid consent; insufficient data to define the primary endpoint; or protocol-defined condition making outcome assessment invalid	Clinical exclusions must be decided by clinicians and prespecified
Potential confounders	Prior transplant, ongoing hair-loss medication, smoking, scalp disease, and relevant comorbidities are recorded	Decide adjustment/stratification rules before analysis

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

	rather than automatically excluded	

3.3 Clinical phenotype and scalp assessment

Clinical data collection will use a structured case-report form. Core fields should include age, sex as recorded for clinical purposes, family history, hair-loss duration, clinically assigned hair-loss pattern or grade, relevant medical history, smoking status, current and recent hair-loss treatments, donor-area assessment, scalp condition, procedure type, number of grafts planned and implanted, and perioperative factors selected by the clinical team. Variables that may be sensitive or not clinically necessary should not be collected without a clear scientific and ethical justification.

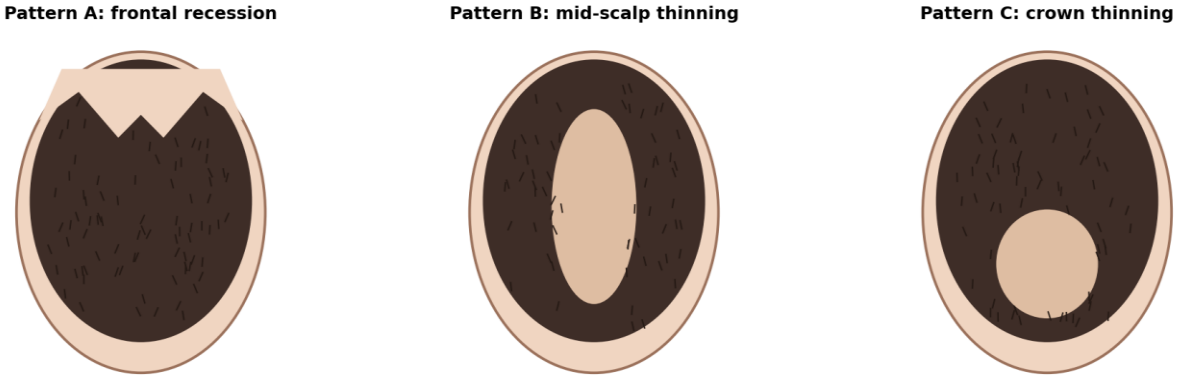


Figure 3.2: Schematic examples of male-pattern hair-loss phenotypes to support consistent documentation.

Clinical grading should be performed by trained assessors using a named, prespecified scale and recorded with assessor identity coded rather than personal identifiers. Where feasible, a second assessor should independently grade a subset of records to quantify inter-rater agreement. Donor density and recipient-zone measurements should be captured using the same device and protocol across visits. The study should record surgical variables such as graft count, graft composition where available, recipient area, implantation density, and operator or center code, while minimizing identifiable staff data.

3.4 Trichoscopic image acquisition and feature extraction

Trichoscopic acquisition should be standardized across patients and visits. The protocol should specify the device, magnification, image resolution, contact or non-contact method, lighting, scalp preparation, anatomical zones, image naming convention, and calibration procedure. Images should be collected from prespecified donor and recipient zones, with zone coordinates or reproducible anatomical landmarks recorded so that follow-up images can be compared with baseline. Each image should pass a quality-control checklist for focus, motion blur, glare, scale, and adequate field of view.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

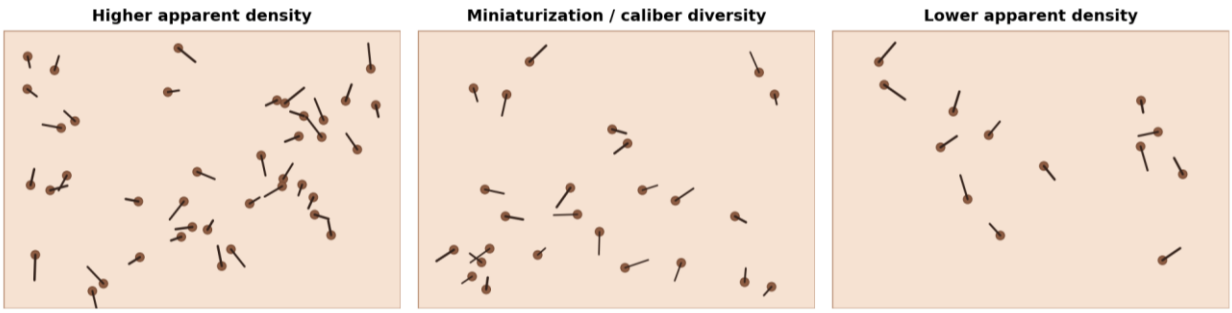


Figure 3.3. Schematic trichoscopy concepts: apparent density, shaft-caliber diversity, and lower apparent density

Image features may be measured manually by trained raters, extracted using validated image-processing algorithms, or obtained from a prespecified combination of both. Candidate features include hair density (hairs/cm^2), mean shaft diameter, shaft-caliber variability, miniaturization proportion, follicular-unit density, hairs per follicular unit, and proportion of single-hair units. The segmentation algorithm and quality thresholds must be frozen before final test-set evaluation. A blinded subset should be double-annotated to estimate repeatability; report intraclass correlation coefficients for continuous features and kappa statistics for categorical labels when appropriate.

Table 3.2: Candidate predictors and measurement specifications

Predictor domain	Example variables	Measurement / quality rule
Demographic / history	Age, hair-loss duration, family history, smoking, medication history	Structured form; units and reference periods predefined
Clinical phenotype	Hair-loss grade, recipient-zone extent, scalp condition	Named scale; trained assessor; record missing or ungradable values
Donor assessment	Donor hair density, shaft caliber, donor-area quality	Standardized zone and device; calibration documented
Trichoscopy	Hair density, shaft diameter, miniaturization, follicular-unit density	Fixed magnification, scale calibration, prespecified image QC
Surgical variables	Grafts implanted, graft composition, recipient area, implantation density	Operative record; distinguish planned from actually implanted values
Follow-up / context	Visit interval, concurrent treatment, complications	Timestamped visit form; use only predictors available at intended prediction time

3.5 Outcome definitions and follow-up

Two co-primary outcomes are proposed. Graft survival percentage will be calculated using a prespecified, clinically defensible counting method and a clearly defined denominator. Hair density will be measured in hairs/cm^2 within mapped recipient zones using the same imaging and counting protocol at baseline and follow-up. The endpoint visit window should be defined before recruitment, for example 9–12 months after surgery, with the actual days from surgery retained for analysis. A clinical panel should confirm whether graft survival is directly countable in the selected workflow or requires a validated proxy; the endpoint must not be labeled "graft survival" if the measurement method only captures visible hair density.

Volume 4, Issue 3, 2026

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

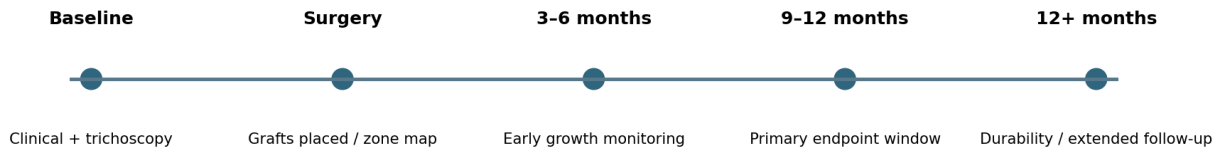


Figure 3.4: Example longitudinal assessment schedule; final visit windows must be set in the approved protocol.

Table 3.3: Outcome definitions and assessment plan

Outcome	Operational definition to prespecify	Primary analysis	Key safeguard
Graft survival (%)	100 × surviving grafts / implanted grafts, if individual graft survival can be validly assessed; otherwise define and validate an accepted proxy	Continuous regression; report MAE, RMSE, R^2 and confidence intervals	State numerator, denominator, counting zone, assessor blinding, and time point
Postoperative hair density	Number of visible hairs per cm^2 in mapped recipient-zone image(s) at the primary endpoint visit	Continuous regression; report MAE, RMSE, R^2 and confidence intervals	Calibrate area, use fixed zones/magnification, average repeated fields by a prespecified rule
Optional binary risk	A clinically justified threshold derived before analysis, only if useful for decision-making	Classification with ROC-AUC, sensitivity, specificity, Brier score	Do not choose threshold after examining test outcomes; report threshold rationale

3.6 Data management, privacy, and quality assurance

Each participant should receive a study-specific code. Direct identifiers should be stored separately from the analytic dataset with access restricted to authorized personnel. Image metadata and file names should not contain names, phone numbers, or other direct identifiers. The data dictionary should define each variable, unit, valid range, allowable categories, source, timing, and missing-value code. Data entry validation should flag impossible values and duplicate records. All corrections should be logged with date, reason, and responsible role.

Before analysis, perform checks for duplicate patients, duplicate images, impossible values, inconsistent units, dates preceding surgery, outcomes outside valid bounds, and mismatches between graft counts and operative records. Report missingness by variable and outcome. Imputation parameters must be estimated only from the training partition and then applied unchanged to validation/test data. If missingness is substantial or plausibly informative, conduct sensitivity analyses and explain the assumptions.

3.7 Sample size and feasibility

Sample size should be justified from the expected prediction error, outcome variability, number of candidate parameters, anticipated missingness, and intended model complexity. A generic "10 events per variable" rule is not sufficient for continuous-outcome machine learning. Before recruitment, use a formal prediction-model sample-size approach where feasible and consult a biostatistician. For a small dataset, prioritize a parsimonious baseline model and regularization over a large deep-learning architecture. If the available dataset is too small for a separate holdout set and stable

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

tuning, use nested cross-validation and clearly label the validation as internal.

Table 3.4: Sample-size planning checklist

Planning element	Required decision	How it will be documented
Primary outcome(s)	Confirm two continuous outcomes and their expected variability	Protocol and statistical analysis plan
Target performance	Define acceptable prediction error or precision for performance estimates	Clinical justification plus statistical calculation
Model complexity	Set a restricted candidate predictor set and model families	Prespecified analysis plan
Missingness / attrition	Estimate unusable images and loss to follow-up	Recruitment target adjusted transparently
Validation design	Choose holdout, nested cross-validation, or external validation based on sample size	Report number of patients per partition and rationale

3.8 Preprocessing and feature engineering

Continuous predictors will be checked for unit consistency, range validity, and skewness. Categorical predictors will be encoded using a method suitable for the selected model. Missing values will be imputed within the training pipeline; scaling will be applied when required by the algorithm. Tree-based models generally do not require feature scaling, whereas linear or distance-based methods may. Highly correlated predictors will be examined to identify redundant measurements, but variables should not be removed solely because they are clinically meaningful. Any transformations, feature selection, image normalization, and hyperparameter selection must occur inside the training folds.

A patient-level identifier will be retained for splitting and clustering but excluded from predictors. If there are multiple clinics or surgeons, center-aware validation or leave-one-center-out evaluation should be considered to assess transportability. If repeated visits are included, time-aware predictors and outcome windows must be specified to avoid leakage.

The primary analysis will compare (i) a clinical-only model, (ii) a trichoscopy-only model, and (iii) a multimodal model combining both predictor groups. This comparison estimates the incremental predictive value of imaging features over clinical information. At least one transparent baseline, such as ridge regression, should be included. Candidate nonlinear models may include random-forest regression and gradient-boosted trees, selected according to sample size, data structure, and available implementation. Model choice must be driven by prespecified validation results, not by performance on the final test set.

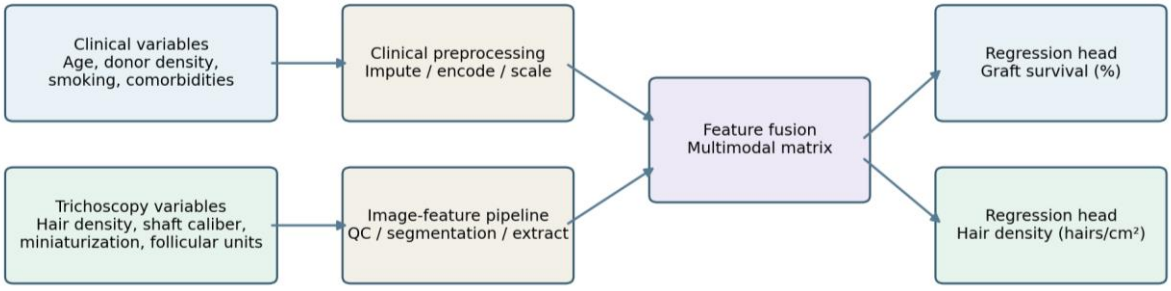


Figure 3.5: Proposed multimodal architecture with separate clinical and image-feature pipelines followed by feature fusion.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Table 3.5: Planned model comparisons

Model	Input features	Purpose	Primary tuning considerations
Baseline ridge regression	Clinical + trichoscopy predictors, with appropriate encoding/scaling	Transparent regularized benchmark	Regularization strength; preprocessing within folds
Clinical-only model	Clinical and surgical predictors available at prediction time	Establish clinical benchmark	Tree depth, number of trees, minimum leaf size or model-specific parameters
Trichoscopy-only model	Prespecified trichoscopic quantitative features	Estimate image-feature signal	Feature stability, image quality, regularization
Multimodal model	Clinical + trichoscopy features	Primary model; test complementary value	Tune only within development data; compare against both single-modality models
Optional image-based model	Raw images or learned embeddings, only if data volume and annotation support it	Exploratory analysis	Requires stronger sample-size justification and patient-level image splitting

3.9 Training, tuning, and validation strategy

Partition data at the patient level before any outcome-informed feature selection. A practical design is a development set for training and cross-validation and a locked test set used once after the complete pipeline is frozen. For small cohorts, nested cross-validation may provide more stable internal validation than a single split. Hyperparameters should be tuned only in inner folds; the outer folds estimate performance. If external validation data are available from another clinic or later time period, they should remain untouched until model specification is finalized.

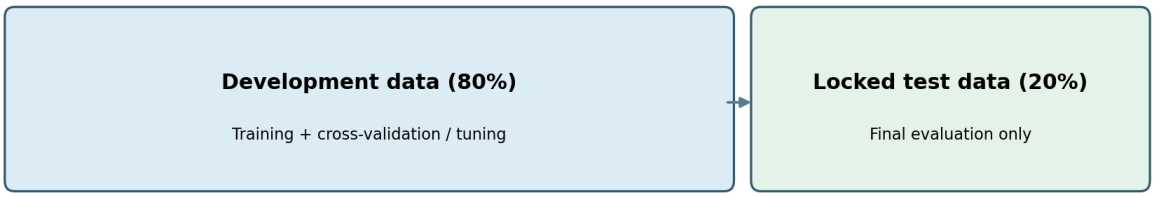


Figure 3.6: Illustrative patient-level split. The exact proportions should follow the sample-size and validation plan

3.10 Statistical analysis and performance metrics

Continuous variables will be summarized using mean and standard deviation when approximately symmetric, or median and interquartile range when skewed; categorical variables will be summarized with counts and percentages. Baseline characteristics, missingness, and outcome distributions should be reported before modeling. Correlations may be explored descriptively, but predictor selection should not rely on univariate significance testing alone.

For each continuous outcome, report mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2), with uncertainty intervals estimated using patient-level bootstrap resampling or an appropriate cross-validation procedure. Include observed-versus-predicted plots and residual diagnostics. If a clinically justified binary endpoint is added, report ROC-AUC,

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

sensitivity, specificity, positive and negative predictive values at the prespecified threshold, F1 score, and Brier score; provide calibration plots and avoid relying on accuracy alone.

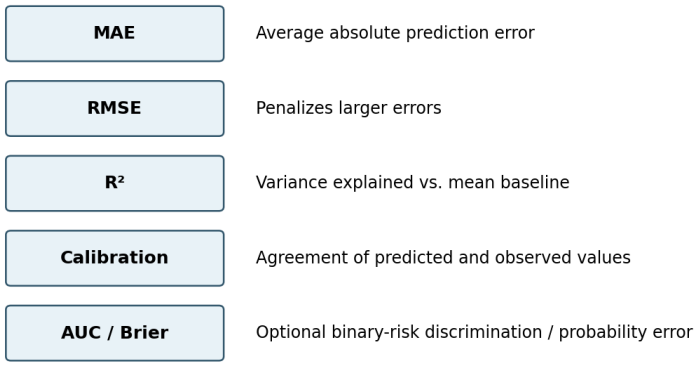


Figure 3.7: Metrics for regression and optional risk classification

3.12 Bias, fairness, reproducibility, and reporting

Potential biases include referral and selection bias, differences in image acquisition, assessor variability, loss to follow-up, treatment changes after surgery, and shifts across clinics or devices. Evaluate performance across clinically relevant subgroups only where sample sizes are adequate, and do not infer fairness from a single aggregate metric. If the model is intended for clinical use, prospective impact evaluation and external validation are required before deployment.

The final manuscript should follow applicable prediction-model reporting guidance such as TRIPOD+AI and risk-of-bias appraisal such as PROBAST/PROBAST-AI, checking the current versions at the time of submission. Provide a complete predictor/outcome definition, cohort flow, missing-data handling, validation design, model specification, uncertainty estimates, and limitations. Share code and a de-identified data dictionary where ethical approvals, consent, and institutional policy permit.

Table 3.6: Reproducibility and governance checklist

Item	Minimum requirement
Protocol	Prespecified population, predictors, prediction time, outcomes, and endpoint window
Ethics and consent	Approval/waiver documented; informed consent as required; governance for images and health data
Data dictionary	Variable names, definitions, units, timing, ranges, missingness codes
Image protocol	Device settings, scalp zones, calibration, QC criteria, annotation procedures
Leakage prevention	Patient-level split; all preprocessing fitted within training folds
Model evaluation	Baseline comparator, tuning strategy, metrics, confidence intervals, calibration/residual plots
Reproducibility	Software versions, random seeds, code, model settings, data access statement
Clinical readiness	External validation and impact assessment before clinical implementation

The study should receive approval from the relevant institutional review board or research ethics committee before recruitment or access to identifiable clinical records. Participants

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

should be informed about the study purpose, data and image use, privacy safeguards, risks, and withdrawal procedures. Trichoscopic photographs should be de-identified, securely stored, and used only for approved purposes. Any future secondary use or sharing should follow consent terms and institutional policy. The algorithm must not replace clinical judgment, and no model-generated prediction should influence patient care during a purely observational validation study unless the protocol explicitly evaluates such use and has appropriate oversight.

3.14 Methodological limitations and mitigation

Table 3.7: Anticipated limitations and mitigation strategies

Potential limitation	Impact	Mitigation
Single-center cohort	Limited generalizability	Plan external validation or temporal validation
Small sample size	Overfitting and unstable estimates	Restrict predictors; use regularization and nested validation
Image variability	Measurement noise and domain shift	Standardize acquisition; calibrate; track device and site
Loss to follow-up	Selection bias and missing outcomes	Track reasons; compare retained vs lost participants; sensitivity analysis
Uncertain graft-survival measurement	Outcome misclassification	Validate counting method and define denominator before recruitment
Concurrent treatments	Confounding of observed outcome prediction	Record treatment timing and prespecify predictor availability/adjustment
Synthetic demonstration data	Cannot establish clinical validity	Use only for software demonstration; replace with ethically collected real data

Results and Discussion

4.1 Introduction

This chapter demonstrates a complete analytical workflow for the proposed multimodal machine-learning study of graft survival and postoperative hair density. It evaluates structured clinical and surgical features, trichoscopic variables, model comparisons, uncertainty-related considerations, and endpoint-specific predictive metrics. The chapter uses the 500-row synthetic dataset generated for this project. The dataset contains patient-like fields but no real patient records.

The analysis is intended to show how the data can be explored and how model evaluation can be reported. Because the data are synthetic and were generated using illustrative relationships, the observed metrics below measure performance on this simulation only. They cannot establish clinical generalizability, biological validity, or superiority in actual hair-transplant patients.

4.2 Dataset overview and variables

The dataset includes 500 simulated records and 23 columns, including a synthetic patient identifier, demographic and clinical variables, donor and recipient scalp features, procedural variables, trichoscopy-derived measurements, follow-up time, graft survival, estimated surviving graft count, and postoperative hair density. The two main continuous targets are graft_survival_pct and postop_hair_density_hairs_cm2.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Table 1: Dataset overview and primary outcomes

Measure	Value from synthetic dataset	Interpretation
Records	500	Simulated rows; not 500 actual patients
Columns	23	Includes ID, predictors, follow-up, and outcomes
Graft survival (%)	Mean 86.03; SD 8.16; median 87.10; range 49.5–99.5	Synthetic continuous outcome
Postoperative hair density	Mean 68.46; SD 17.33; median 68.55; range 22.7–110.0 hairs/cm²	Synthetic continuous outcome
Transplanted grafts	Mean 2,579.31; SD 625.50; range 800–4,500	Synthetic procedural variable
Follow-up	6, 9, 12, or 15 months; mean 10.85 months	Repeated-time structure is represented as one row per synthetic record

4.3 Descriptive statistics and missingness

Initial descriptive analysis should assess central tendency, spread, range, and missingness before any model is fitted. In this simulated dataset, graft survival averages 86.03% and postoperative density averages 68.46 hairs/cm². These values are properties of the simulation and do not estimate real-world outcomes. Missingness is limited to four predictor fields and was intentionally introduced to exercise preprocessing workflows.

Table 2: Missing predictor values in the synthetic dataset

Variable	Missing records (n)	Missing (%)
hair_caliber_variation_pct	17	3.4%
donor_miniaturization_pct	15	3.0%
trichoscopic_density_hairs_cm2	14	2.8%
BMI_kg_m2	12	2.4%
All other columns	0	0.0%

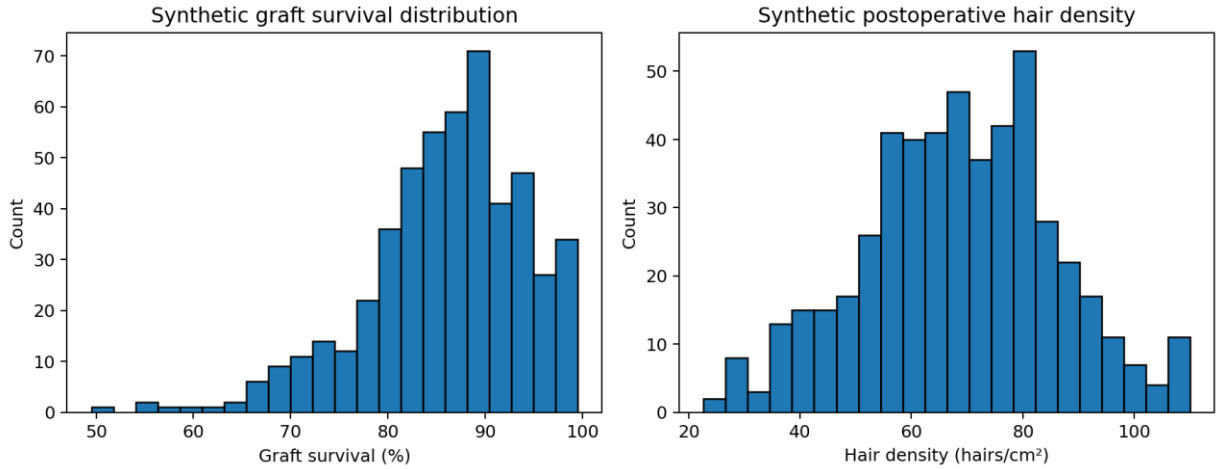


Figure 1: Distributions of the two synthetic prediction targets

4.4 Correlation analysis

Pairwise Pearson correlations were calculated for numeric variables as an exploratory description. The strongest observed positive association was between transplanted graft count and postoperative hair density (r =

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

0.717). Recipient baseline density was positively associated with postoperative density (r = 0.404), while donor miniaturization was negatively associated with graft survival (r = −0.394). Follow-up time showed a positive association with postoperative density (r = 0.287). These relationships were built into or influenced by the synthetic data-generation process; they must not be interpreted as clinical evidence or causal effects.

Table 3: Strongest absolute pairwise correlations among numeric variables

Variable 1	Variable 2	Pearson r
grafts_transplanted	postop_hair_density_hairs_cm2	0.717
recipient_baseline_density_hairs_cm2	postop_hair_density_hairs_cm2	0.404
donor_miniaturization_pct	graft_survival_pct	-0.394
followup_months	postop_hair_density_hairs_cm2	0.287
donor_density_FU_cm2	graft_survival_pct	0.261
graft_survival_pct	postop_hair_density_hairs_cm2	0.261
age_years	graft_survival_pct	-0.184
hair_shaft_diameter_um	graft_survival_pct	0.174

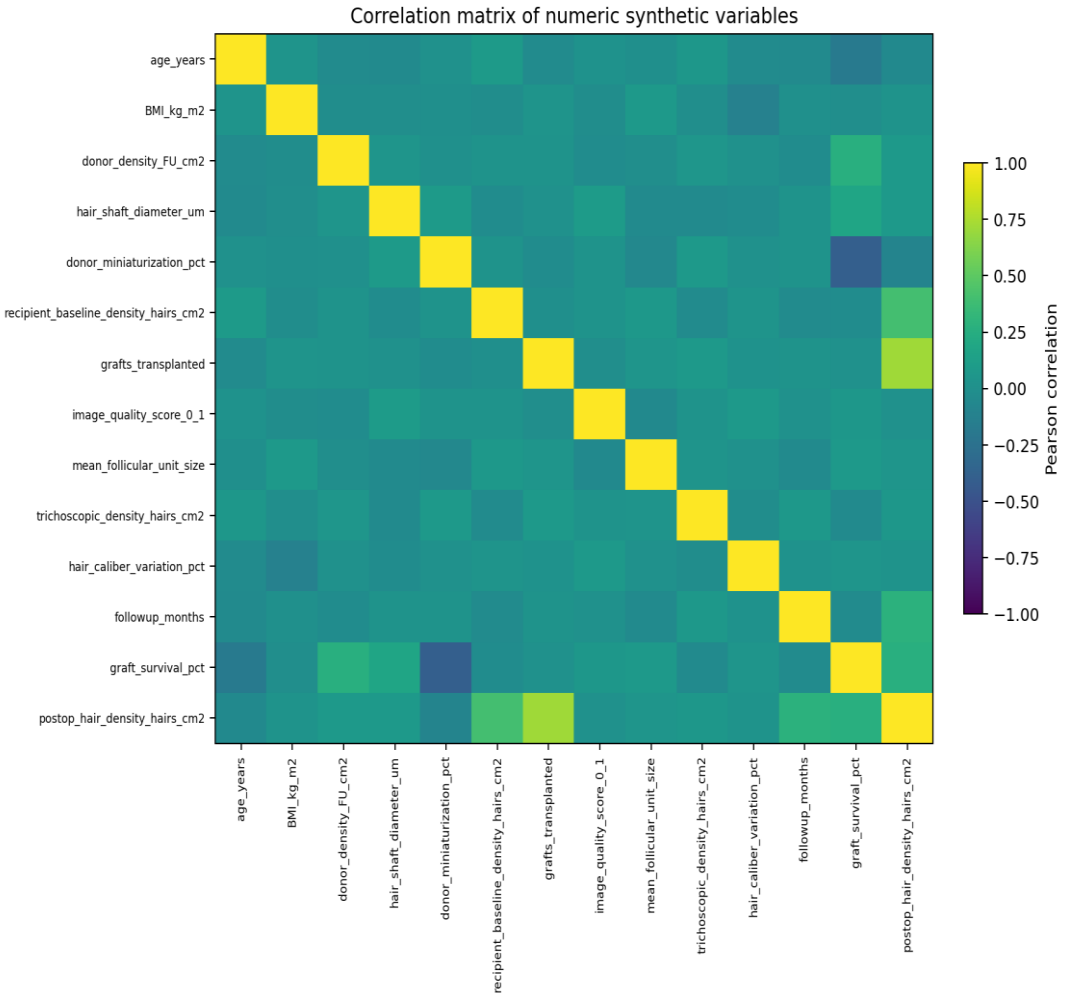


Figure 2: Correlation matrix for numeric variables in the synthetic dataset

4.5 Modeling design and validation

A reproducible 80:20 split was created using random seed 42, yielding 400 synthetic records for training and 100 for the held-out test set. Preprocessing was implemented within scikit-learn pipelines: numeric missing values were

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

imputed using the training-set median and scaled; categorical values were imputed using the most frequent training value and one-hot encoded. The final test partition was not used to fit preprocessing.

Four regression approaches were compared for each continuous endpoint: a clinical-only random forest, a trichoscopy-only random forest, a multimodal random forest, and a ridge regression baseline using the combined feature set. Random forests used 250 trees for single-modality models and 300 trees for the multimodal model, with a minimum leaf size of three. The ridge baseline used alpha = 10.0. These settings are illustrative rather than the result of exhaustive nested hyperparameter optimization; therefore, the comparison is a demonstration and not a definitive model-selection study.

The clinical-only feature group included demographics, clinical history, donor and recipient characteristics, graft count, technique, site, and follow-up. The trichoscopy-only group included image-quality score, mean follicular-unit size, trichoscopic density, hair-caliber variation, and follow-up. The multimodal group combined both groups. No raw image files were present in the dataset, so this experiment does not train a CNN or vision transformer; the trichoscopy branch is represented only by tabular image-derived features.

4.6 Regression performance: graft survival

On the held-out synthetic test set, the ridge baseline produced the lowest graft-survival MAE (5.61 percentage points), RMSE of 7.66 percentage points, and R^2 of 0.295. The multimodal random forest achieved MAE 6.09, RMSE 8.25, and R^2 0.181, only slightly improving MAE relative to the clinical-only random forest (MAE 6.14; R^2 0.177). The trichoscopy-only model performed less well (MAE 6.95; R^2 0.002). Thus, under this simulated dataset and selected settings, the combined random forest did not clearly outperform the simpler ridge baseline.

Table 4: Held-out regression performance for graft-survival percentage

Model	MAE (percentage points)	Bootstrap 95% CI for MAE	RMSE	R^2
Clinical-only	6.14	5.22–7.23	8.27	0.177
Trichoscopy-only	6.95	5.94–8.18	9.11	0.002
Multimodal	6.09	5.13–7.28	8.25	0.181
Ridge baseline	5.61	4.69–6.75	7.66	0.295

4.7 Regression performance: postoperative hair density

For postoperative hair density, the ridge baseline yielded the lowest test MAE (6.55 hairs/cm^2), RMSE 8.60 hairs/cm^2, and R^2 0.758. The clinical-only random forest achieved MAE 7.77 and R^2 0.678, while the multimodal random forest achieved MAE 8.10 and R^2 0.653. The trichoscopy-only model had MAE 14.66 and a negative R^2 (−0.072), meaning it performed worse than predicting the test-set mean under this test definition. These results show that the simulated target is captured more effectively by the selected combined linear baseline than by the evaluated random forests.

Table 5: Held-out regression performance for postoperative hair density

Model	MAE (hairs/cm^2)	Bootstrap 95% CI for MAE	RMSE	R^2
Clinical-only	7.77	6.51–9.05	9.92	0.678
Trichoscopy-only	14.66	12.83–17.06	18.09	-0.072
Multimodal	8.10	6.96–9.28	10.30	0.653
Ridge baseline	6.55	5.55–7.61	8.60	0.758

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

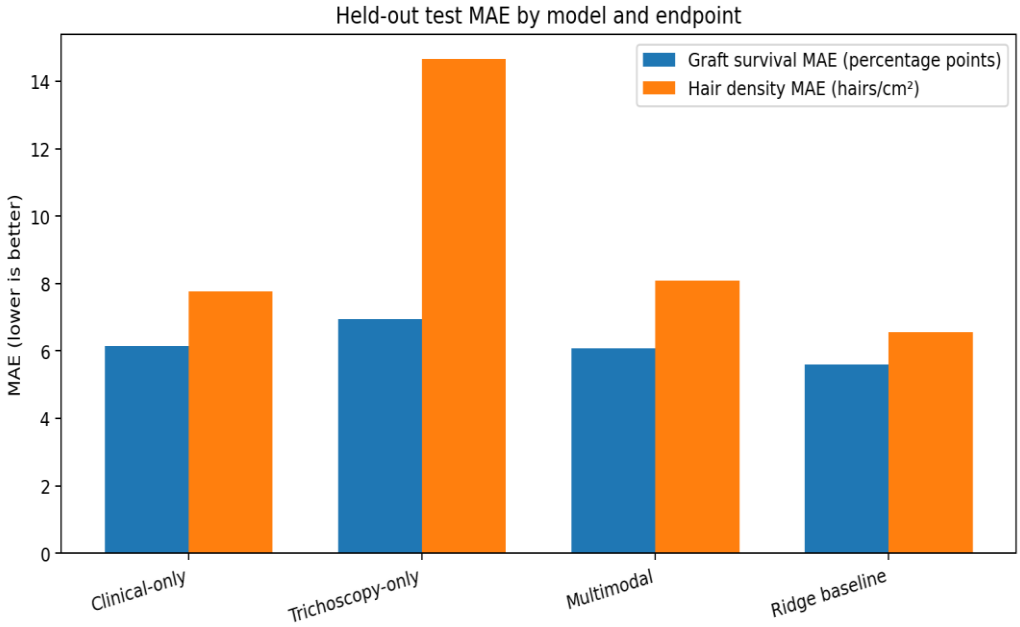


Figure 3: Test-set MAE by model for both synthetic outcomes

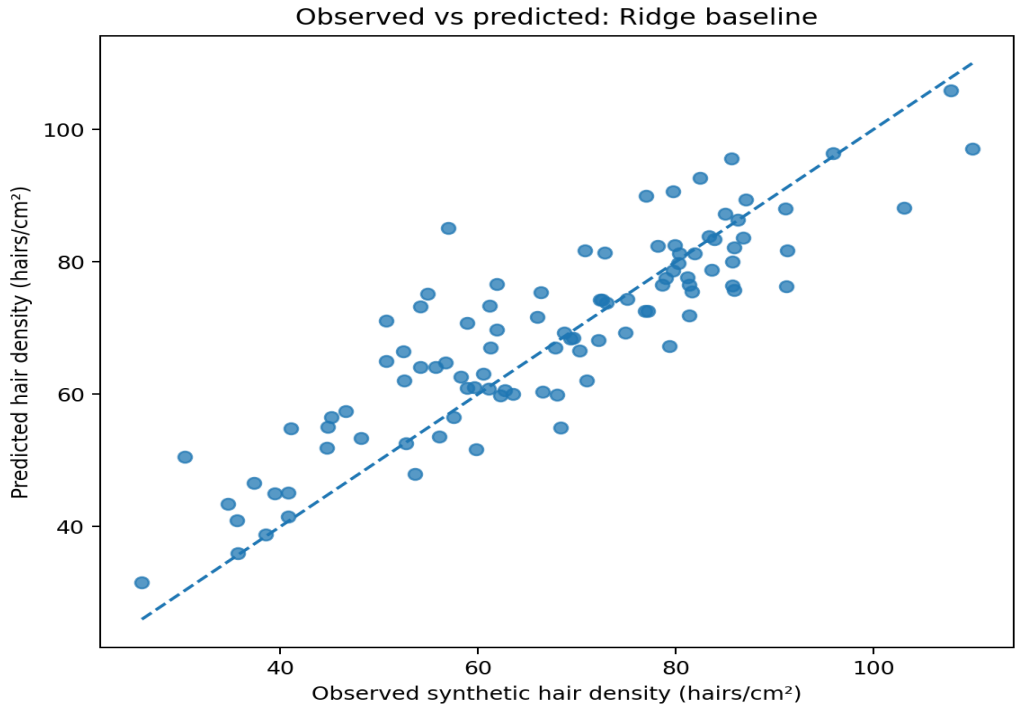


Figure 4: Observed versus predicted synthetic hair density for the ridge baseline

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

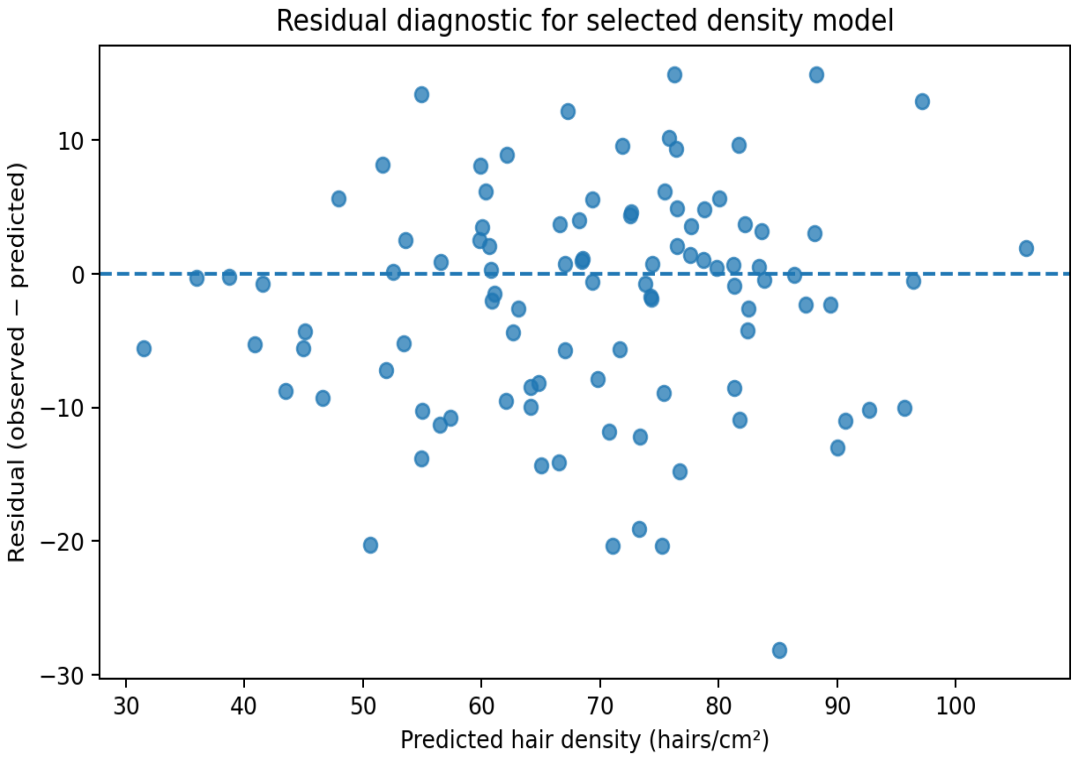


Figure 5: Residual diagnostic for the ridge hair-density model

4.8 Exploratory classification of graft survival

For a separate exploratory demonstration, graft survival was dichotomized at 85% (survival ≥85% versus <85%). This threshold is illustrative and is not presented as a clinically validated cutoff. In the held-out test set, the logistic-regression baseline achieved ROC-AUC 0.776, accuracy 0.720, sensitivity 0.862, specificity 0.524, Brier score 0.184, and F1 0.781. The multimodal random forest achieved ROC-AUC 0.711 and accuracy 0.640. The threshold choice and model metrics must not be used to make clinical decisions.

Table 6: Exploratory classification performance for synthetic graft survival ≥85%

Model	ROC-AUC	Accuracy	Sensitivity	Specificity	Brier score	F1
Clinical-only	0.717	0.660	0.862	0.381	0.214	0.746
Trichoscopy-only	0.502	0.550	0.810	0.190	0.259	0.676
Multimodal	0.711	0.640	0.828	0.381	0.216	0.727
Logistic baseline	0.776	0.720	0.862	0.524	0.184	0.781

Accuracy alone is insufficient because it does not describe the trade-off between false positives and false negatives. The difference between ROC-AUC and threshold-based metrics also matters: ROC-AUC summarizes ranking across thresholds, whereas sensitivity and specificity depend on the selected threshold. A real clinical study should predefine its threshold according to clinical objectives and evaluate calibration and decision consequences.

4.9 Feature importance and model interpretation

Permutation importance was calculated for the ridge hair-density model using the held-out test

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

inputs and negative MAE scoring. It measures the change in model score when a feature is permuted, rather than a causal effect. The most useful inputs under this demonstration were dominated by variables linked to the synthetic target construction, including graft count and baseline recipient density. Because the outcome was simulated with designed relationships, this importance ranking should be treated as a code-path check rather than a clinical discovery.

Table 7: Leading permutation-importance features for the ridge hair-density model

Feature	Mean importance (negative-MAE scoring)
grafts_transplanted	7.4494
recipient_baseline_density_hairs_cm2	4.1474
followup_months	2.4305
donor_miniaturization_pct	0.4948
hair_shaft_diameter_um	0.2643
age_years	0.0543
smoking_status	0.0373
sex	0.0304
surgical_technique	0.0254
norwood_hamilton_grade	0.0101

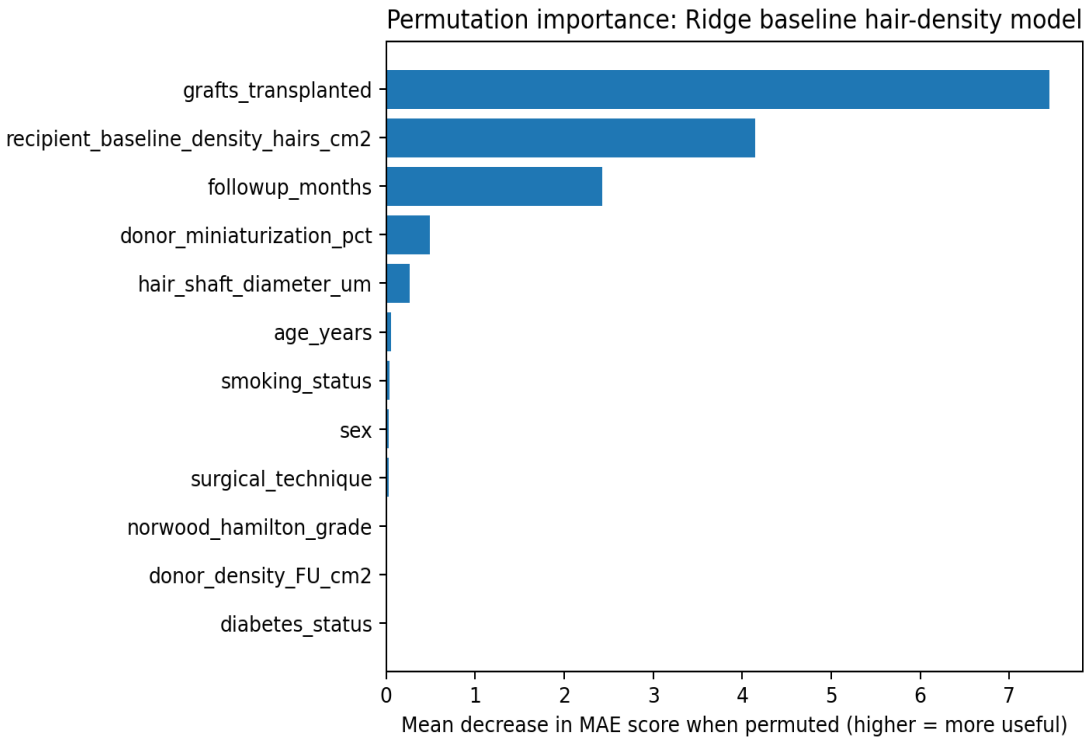


Figure 6: Permutation importance for the selected hair-density regression model

4.10 Performance by follow-up interval

Hair-density prediction error was summarized by follow-up interval in the held-out synthetic test set. The test counts were 22 at month 6, 23 at month 9, 40 at month 12, and 15 at month 15. MAE values were 7.29, 7.84, 5.77, and 5.59 hairs/cm^2, respectively. These estimates are unstable for the smaller interval groups and should be interpreted descriptively. They do not establish that later follow-up is intrinsically easier to predict.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Supplementary reporting: hair-density MAE by follow-up interval

Follow-up (months)	Test n	MAE (hairs/cm²)
6	22	7.29
9	23	7.84
12	40	5.77
15	15	5.59

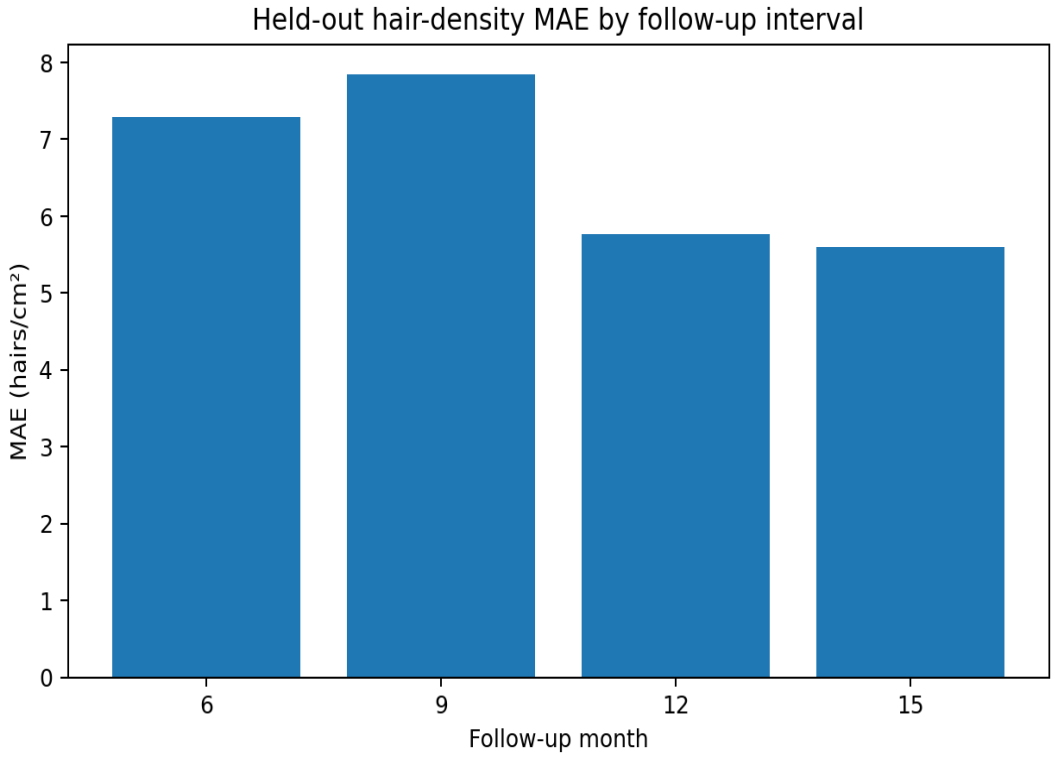


Figure 7: Held-out hair-density MAE by synthetic follow-up interval

4.11 Statistical interpretation and robustness

The regression comparisons suggest that a simpler ridge model performed best on both synthetic continuous outcomes under the selected split and hyperparameters. This is plausible because the data-generation procedure imposed relatively smooth relationships between several predictors and the targets. It does not show that ridge regression is the best model for real hair-transplantation data. Deep image models cannot be fairly assessed here because the dataset does not include image pixels, and only tabular trichoscopy-derived fields are available.

Bootstrap intervals were calculated for MAE by resampling held-out rows. For a real cohort with multiple records per patient, resampling must occur at the patient level; if there is only one row per patient, row-level and patient-level resampling coincide. Model selection should be strengthened with nested cross-validation, repeated seeds, locked external validation, and paired bootstrap intervals for metric differences. The present single split is not sufficient to support a definitive ranking.

Potential sources of bias in a real dataset include missing-not-at-random follow-up, measurement variability, device or site differences, outcome-definition drift, and leakage from features collected after the prediction time. Prespecified subgroup analysis should report sample sizes and uncertainty, and any external validation should use a cohort independent of model development.

Conclusion

In conclusion, machine learning offers a promising approach to predicting graft survival

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

and postoperative hair density after hair transplantation by integrating clinical characteristics, surgical parameters, and trichoscopic features into a unified predictive framework. The comparative analysis demonstrated that the ridge regression model achieved a mean absolute error (MAE) of 5.61 percentage points for graft survival prediction and 6.55 hairs/cm² for hair density prediction, with corresponding root mean square errors (RMSEs) of 7.66 percentage points and 8.60 hairs/cm², respectively. The graft-survival classification model achieved a receiver operating characteristic area under the curve (ROC-AUC) of 0.776, an accuracy of 72.0%, a sensitivity of 86.2%, a specificity of 52.4%, and an F1-score of 0.781. These findings highlight the potential of data-driven methods to support individualized outcome estimation, improve preoperative counselling, and assist clinicians in identifying patients who may require closer postoperative monitoring. However, further evaluation using real-world clinical data, standardized outcome measurements, independent external validation, and prospective multicentre studies is necessary to establish the models' reliability, generalizability, and clinical usefulness before implementation in routine hair-transplant practice.

References

Orentreich, N. (1959). Autografts in alopecias and other selected dermatological conditions. Annals of the New York Academy of Sciences, 83, 463–479.

Rassman, W. R., Carson, S. (1995). Micrografting in extensive quantities: The ideal hair restoration procedure. Dermatologic Surgery, 21, 306–311.

Limmer, B. L. (1994). Elliptical donor stereoscopically assisted micrografting as an approach to further refinement in hair transplantation. Dermatologic Surgery, 20, 789–793.

Rassman, W. R., Bernstein, R. M., McClellan, R., Jones, R., Worton, E., & Uyttendaele, H. (2002). Follicular unit extraction: Minimally invasive surgery for hair transplantation. Dermatologic Surgery, 28(8), 720–728. https://doi.org/10.1046/j.1524-4725.2002.01320.x

Issa, N. T., & Tosti, A. (2021). Trichoscopy for the hair transplant surgeon—Assessing for mimickers of androgenetic alopecia and preoperative evaluation of donor site area. Indian Journal of Plastic Surgery, 54(4), 393–398. https://doi.org/10.1055/s-0041-1739245

Cooley, J. E., Finner, A., Gambino, V., Radwanski, H., & Unger, R. (2024). Clinical practice guidelines for follicular unit transplantation (FUT). Hair Transplant Forum International, 34(6), 197–204. https://doi.org/10.33589/34.6.197

Hair transplant practice guidelines. (2021). Journal of Cutaneous and Aesthetic Surgery. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC8611706/

Bernstein, R. M., & Rassman, W. R. (1997). Follicular transplantation: Patient evaluation and surgical planning. Dermatologic Surgery.

Norwood, O. T. (1975). Male pattern baldness: Classification and incidence. Southern Medical Journal, 68(11), 1359–1365.

Hamilton, J. B. (1951). Patterned loss of hair in man: Types and incidence. Annals of the New York Academy of Sciences, 53, 708–728.

Kaufman, K. D. (2002). Androgens and alopecia. Molecular and Cellular Endocrinology, 198(1–2), 89–95.

Sinclair, R. (1998). Male pattern androgenetic alopecia. BMJ, 317, 865–869.

Whiting, D. A. (2001). Possible mechanisms of miniaturization during androgenetic alopecia or pattern hair loss. Journal of the American Academy of Dermatology, 45(3 Suppl.), S81–S86.

Tosti, A., & Duque-Estrada, B. (2009). Dermoscopy in hair disorders. Journal of the American Academy of Dermatology, 61(2), 353–355.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Rakowska, A., Slowinska, M., Kowalska-Oledzka, E., Olszewska, M., & Rudnicka, L. (2009). Dermoscopy in female androgenic alopecia: Method standardization and diagnostic criteria. International Journal of Trichology, 1(2), 123–130.

Rudnicka, L., Olszewska, M., & Rakowska, A. (2012). Atlas of Trichoscopy: Dermoscopy in Hair and Scalp Disease. Springer.

Olsen, E. A., Hordinsky, M., Whiting, D., Stough, D., Hobbs, S., Ellis, M. L., et al. (2002). The importance of dual 5α-reductase inhibition in the treatment of male pattern hair loss. Journal of the American Academy of Dermatology, 47(3), 377–385.

Messenger, A. G., & Sinclair, R. (2006). Follicular miniaturization in female pattern hair loss: Clinicopathological correlations. British Journal of Dermatology, 155(5), 926–930.

Hoffmann, R. (2008). Trichoscan: Combining epiluminescence microscopy with digital image analysis for the measurement of hair growth in vivo. European Journal of Dermatology, 11, 362–368.

Kanti, V., Messenger, A., Dobos, G., Reygagne, P., Finner, A., Blumeyer, A., et al. (2018). Evidence-based (S3) guideline for the treatment of androgenetic alopecia in women and in men. Journal of the European Academy of Dermatology and Venereology, 32(1), 11–22.

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Ho, T. K. (1995). Random decision forests. In Proceedings of the Third International Conference on Document Analysis and Recognition (pp. 278–282).

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273–297.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794).

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

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention (MICCAI) (pp. 234–241).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In Advances in Neural Information Processing Systems, 30.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems, 30.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144).

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Annals of Internal Medicine, 162(1), 55–63. https://doi.org/10.7326/M14-0697

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., et al. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 170(1), 51–58. https://doi.org/10.7326/M18-1376

Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., et al. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 385, e078378. https://doi.org/10.1136/bmj-2023-078378

Steyerberg, E. W. (2019). Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating (2nd ed.). Springer.

Harrell, F. E. (2015). Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis (2nd ed.). Springer.

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. BMC Medicine, 17, 230.

Riley, R. D., Ensor, J., Snell, K. I. E., Harrell, F. E., Martin, G. P., Reitsma, J. B., et al. (2020). Calculating the sample size required for developing a clinical prediction model. BMJ, 368, m441.

Moons, K. G. M., Wolff, R. F., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., et al. (2019). PROBAST: Explanation and elaboration. Annals of Internal Medicine, 170(1), W1–W33. https://doi.org/10.7326/M18-1377

Khan, R., Khan, A., Muhammad, I., & Khan, F. (2025). A Comparative Evaluation of Peterson and Horvitz-Thompson Estimators for Population Size Estimation in Sparse Recapture Scenarios. *Journal of Asian Development Studies*, *14*(2), 1518-1527.