

TRIPOD+AI reporting checklist

Dermoscopy-Dominant Correction Fusion for Paired Clinical and Dermoscopic Skin Lesion Image Classification

Yusung Chu and Sejung Yang | Diagnostic model development and internal evaluation on MILK10k

Reporting basis: Manuscript, Supplementary Material, and the MILK10k dataset report [2]. Locations are given by section title/number and table/figure identifier so that they remain valid after typesetting.

Dataset [2] denotes source collection/curation information, not work performed by the Dermoscopy-Dominant Correction Fusion (DDCF) investigators. Source demographics refer to the original 5,240-lesion cohort, not the 5,187-lesion DDCF analytical subset.

Status labels: Reported; Partly reported; Not assessed; Not applicable; Not available; Not explicitly reported. “Not assessed” means the analysis was not part of the reported DDCF evaluation; “Not available” means the source report/data do not provide the requested information. Labels describe reporting status, not study quality.

Main checklist — Title, abstract and background

27 main items / 52 subitems. D = model development; E = model evaluation. Item wording follows TRIPOD+AI [1].

Item	Reporting requirement	Section / supporting source	Reporting information
1 D / E	Title Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	Title; Abstract — Background and Methods	[Partly reported] The title identifies DDCF, paired clinical and dermoscopic images, and skin lesion classification. Model development and retrospective internal evaluation are stated in the abstract and Methods; these terms are not explicit in the title.
2 D / E	Abstract See TRIPOD+AI for Abstracts checklist	Abstract — Background, Methods, Results and Conclusions	[Reported] The structured abstract summarizes the diagnostic model, fixed-coefficient experiments, lowest-validation-loss checkpoint selection, three internal partitions, performance metrics and principal results.
3a D / E	Background Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model, including references to existing models	Background; Methods 2.1	[Reported] The diagnostic context and complementary information in clinical and dermoscopic images motivate asymmetric fusion. Existing multimodal approaches are cited, and DDCF is introduced as a dermoscopic base representation with gated clinical correction.
3b D / E	Background Describe the target population and the intended purpose of the prediction model in the context of the care pathway, including its intended users (eg, healthcare professionals, patients, public)	Background; Methods 2.1; Conclusions Dataset [2]: Data collection and selection; Usage notes	[Partly reported] The model is evaluated for diagnosis of lesions with paired clinical and dermoscopic images. The source population mainly comprises lesions assessed when skin cancer was in the differential diagnosis, rather than a general-population screening sample [2]. The current study is a retrospective computational evaluation; a deployed clinical workflow and intended end-user role are not specified.
3c D / E	Background Describe any known health inequalities between sociodemographic groups	Dataset [2]: Skin tone; Usage notes Discussion — limitations	[Partly reported] The dataset report identifies underrepresentation of darker skin tones and children, and limited coverage of some anatomical sites [2]. These are source-data representation limitations relevant to equitable applicability; they are not evidence that DDCF was tested for demographic fairness. The DDCF manuscript does not separately characterize broader healthcare inequalities.
4 D / E	Objectives Specify the study objectives, including whether the study describes the development or validation of a prediction model (or both)	Abstract; Background — study objective; Methods 2.1	[Reported] The objective is development and internal evaluation of DDCF to assess whether modality-asymmetric fusion improves paired clinical–dermoscopic classification. Three backbone families and repeated lesion-level partitions are evaluated; no external-validation study is reported.

Data source, setting and eligibility

Item	Reporting requirement	Section / supporting source	Reporting information
5a D / E	Data Describe the sources of data separately for the development and evaluation datasets (eg, randomised trial, cohort, routine care or registry data), the rationale for using these data, and representativeness of the data	Methods 2.1; Background Dataset [2]: Ethics review; Data collection and selection; Usage notes	[Partly reported] Training, validation and test subsets come from the same eligible public MILK10k cohort. The source images were collected retrospectively during routine care at five centres. The paired imaging design supports the fusion comparison. Biopsy/excision-based selection and underrepresentation of clinically unremarkable benign lesions restrict representativeness [2].
5b D / E	Data Specify the dates of the collected participant data, including start and end of participant accrual; and, if applicable, end of follow-up	Dataset [2]: Data collection and selection Methods 2.1	[Not available] The source report states that collection was continuous in Vienna and Queensland and used convenience sampling at the other institutions [2], but it does not provide calendar start/end dates of participant/image accrual. No longitudinal follow-up applies to this diagnostic classification outcome.
6a D / E	Participants Specify key elements of the study setting (eg, primary care, secondary care, general population) including the number and location of centres	Methods 2.1 Dataset [2]: Ethics review	[Partly reported] The five source settings comprise three tertiary referral centres (Medical University of Vienna, Ankara University and Memorial Sloan Kettering Cancer Center), a dermatology practice in Skopje, and a general medical practice with a special interest in skin cancer in Capalaba, Australia [2]. This specifies the care settings behind the countries reported in Methods.
6b D / E	Participants Describe the eligibility criteria for study participants	Methods 2.1; Table 1 Dataset [2]: Data collection and selection; Table 1	[Partly reported] The source included high-quality paired images of pigmented/nonpigmented lesions considered for biopsy or excision, plus a limited number of straightforward clinically diagnosed lesions; nail pigmentations were excluded [2]. DDCF retained nine diagnostic categories, excluding 44 other-benign and 9 other-malignant lesions, leaving 5,187 paired lesions.
6c D / E	Participants Give details of any treatments received, and how they were handled during model development or evaluation, if relevant	Methods 2.1 Dataset [2]: Data collection and selection	[Not applicable] Not applicable to the reported DDCF objective: the model predicts the current lesion diagnosis, not treatment response or a future clinical outcome. Biopsy/excision serves as reference-standard ascertainment in the source dataset [2], not as a treatment predictor.
7 D / E	Data preparation Describe any data pre-processing and quality checking, including whether this was similar across relevant sociodemographic groups	Methods 2.1, 2.3 and 2.4 Dataset [2]: Data collection and selection; Data processing	[Partly reported] Source curation excluded poor-quality/identifiable images and ambiguous pathology; removed EXIF data; standardized image format, aspect ratio, orientation and resolution; cropped/blurred identifying features; and removed suspected duplicates/label errors [2]. DDCF used 224×224 inputs; FDDM/CRD-Net used 448×448 . DDCF-specific normalization, augmentation and demographic-specific quality checks are not detailed.

Outcome ascertainment, predictors and sample size

Item	Reporting requirement	Section / supporting source	Reporting information
8a D / E	Outcome Clearly define the outcome that is being predicted and the time horizon, including how and when assessed, the rationale for choosing this outcome, and whether the method of outcome assessment is consistent across sociodemographic groups	Methods 2.1; Table 1 Dataset [2]: Data collection and selection; Dataset description; Table 2	[Partly reported] The outcome is the supplied nine-category lesion diagnosis, with no prognostic time horizon. In the full 5,240-lesion source cohort, 5,016 (95.7%) diagnoses used histopathology; non-biopsied straightforward lesions used treating-physician diagnosis confirmed by two authors [2]. This percentage is not a new estimate for the 5,187-lesion DDCF subset. Assessment consistency by demographic group is not reported.
8b D / E	Outcome If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of the outcome assessors	Dataset [2]: Data collection and selection; author affiliations	[Partly reported] Original pathology diagnoses were obtained from participating centres. Non-biopsied clinical diagnoses were confirmed by Philipp Tschandl and Harald Kittler; these two authors also mapped labels to diagnostic taxonomies and are affiliated with dermatology [2]. Individual pathologists' credentials and demographic characteristics of all diagnostic assessors are not reported.
8c D / E	Outcome Report any actions to blind assessment of the outcome to be predicted	Dataset [2]: Data collection and selection Methods 2.1 and 2.3	[Not available] Formal blinding of the original diagnostic outcome assessment is not described in the MILK10k source report [2]. DDCF used the released diagnostic labels; validation/test separation is a model-evaluation procedure and is not treated as blinding of the original outcome assessment.
9a D	Predictors Describe the choice of initial predictors (eg, literature, previous models, all available predictors) and any pre-selection of predictors before model building	Background; Methods 2.1 and 2.2	[Reported] The two predictors are the clinical close-up and dermoscopic image of the same lesion, selected because of complementary diagnostic content. Neural encoders learn image features; the reported model does not add age, sex, skin tone, anatomical site or dataset concept scores as predictors.
9b D / E	Predictors Clearly define all predictors, including how and when they were measured (and any actions to blind assessment of predictors for the outcome and other predictors)	Methods 2.1–2.4 Dataset [2]: Data collection and selection; Dataset description	[Partly reported] Each released lesion has one clinical close-up and one corresponding dermoscopic image. Clinical photographs used different cameras and non-standardized acquisition protocols across centres; dermoscopy used polarized light [2]. The manuscript specifies modality encoders and input sizes. Exact image-to-diagnosis timing and acquisition/selection blinding are not reported.
9c D / E	Predictors If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of the predictor assessors	Methods 2.1 and 2.2 Dataset [2]: Data collection and selection; Skin tone	[Not applicable] Not applicable to the DDCF predictor definition because the model uses raw clinical and dermoscopic images rather than manually scored predictor features. Source image acquisition was performed by a professional photographer in Vienna and by treating physicians/medical personnel elsewhere [2].
10 D / E	Sample size Explain how the study size was arrived at (separately for development and evaluation), and justify that the study size was sufficient to answer the research question. Include details of any sample size calculation	Methods 2.1 and 2.5; Table 1; Discussion — limitations Dataset [2]: Table 1	[Partly reported] Study size follows the available eligible cohort: 5,240 source lesions minus 44 other-benign and 9 other-malignant lesions = 5,187. Each partition contains 3,630 training, 519 validation and 1,038 test lesions. No formal sample-size or power calculation was performed; the small rare-class groups and associated uncertainty are acknowledged.

Data handling and analytical methods

Item	Reporting requirement	Section / supporting source	Reporting information
11 D / E	Missing data Describe how missing data were handled. Provide reasons for omitting any data	Methods 2.1 Dataset [2]: Dataset description; Table 1	[Partly reported] The source release is described as one image pair and diagnosis per lesion. Its Table 1 reports missing anatomical site for 31/5,240 lesions (0.6%) [2]; anatomical-site metadata is not a reported DDCF input. Diagnostic-category exclusions are documented. Additional DDCF handling of missing/corrupt images or labels, and any imputation, is not explicitly specified.
12a D	Analytical methods Describe how the data were used (eg, for development and evaluation of model performance) in the analysis, including whether the data were partitioned, considering any sample size requirements	Methods 2.1, 2.3 and 2.5; Table 1	[Reported] Three data-splitting seeds generate stratified lesion-level 7:1:2 partitions with identical class-wise counts but different lesion assignments. Training fits weights; lowest validation loss selects the checkpoint within each fixed configuration; held-out test subsets evaluate performance. Paired images stay together. These internal partitions are not independent external cohorts.
12b D	Analytical methods Depending on the type of model, describe how predictors were handled in the analyses (functional form, rescaling, transformation, or any standardisation)	Methods 2.2–2.4 Dataset [2]: Data processing	[Partly reported] The released images underwent source-level standardization [2]. DDCF subsequently uses model-specific resizing, separate encoders/adaptors and 512-dimensional feature projections, followed by gated residual fusion. Source curation is distinguished from network-level preprocessing; the manuscript does not detail input scaling, normalization or augmentation parameters.
12c D	Analytical methods Specify the type of model, rationale†, all model building steps, including any hyperparameter tuning, and method for internal validation	Methods 2.2–2.5; Figure 1; Table 3; Discussion — coefficient analysis	[Partly reported] The manuscript describes DDCF architecture, optimizer, learning rate, batch size, training duration and fixed λ values. Each configuration uses lowest-validation-loss checkpoint selection. The λ sweep is descriptive; $\lambda = 0.7$ is not claimed to be validation-selected. The exact DDCF loss formulation is not stated; the dataset paper's separate training pipeline is not substituted for it.
12d D / E	Analytical methods Describe if and how any heterogeneity in estimates of model parameter values and model performance was handled and quantified across clusters (eg, hospitals, countries). See TRIPOD-Cluster for additional considerations‡	Methods 2.1 and 2.5; Discussion — limitations Dataset [2]: Usage notes	[Not assessed] Centre/country heterogeneity was not assessed in the reported DDCF evaluation. Performance was examined across repeated data partitions rather than source-centre clusters. Individual-image source identifiers are withheld in the source dataset for privacy [2].
12e D / E	Analytical methods Specify all measures and plots used (and their rationale) to evaluate model performance (eg, discrimination, calibration, clinical utility) and, if relevant, to compare multiple models	Methods 2.5; Results 3.1–3.6; Tables 2–5; Figures 2–3; Supplementary Tables S1 and S3	[Partly reported] Primary macro-OvR AUROC and complementary BACC/macro-F1 are defined. Class-wise sensitivity, specificity, F1 and AUROC, paired bootstrap differences, case transitions and uncertainty-stratified displays are reported. The denominators for corrected/adverse proportions are explicit. Calibration and decision-analytic clinical-utility results are not included in the reported evaluation.

Model evaluation, fairness and output

Item	Reporting requirement	Section / supporting source	Reporting information
12f E	Analytical methods Describe any model updating (eg, recalibration) arising from the model evaluation, either overall or for particular sociodemographic groups or settings	Methods 2.3; Results — fixed-configuration evaluation	[Not assessed] No post-evaluation recalibration, model updating, or demographic-specific updating was assessed in the reported DDCF evaluation. The reported configurations are separately trained models selected by validation loss, not post-evaluation updates.
12g E	Analytical methods For model evaluation, describe how the model predictions were calculated (eg, formula, code, object, application programming interface)	Methods 2.2–2.5; Figure 1; Availability of data and materials	[Partly reported] Fusion equations, encoders, validation-loss checkpoint selection and probability-based/top-1 evaluation are described. Code and analysis scripts are available on reasonable request. The manuscript does not specify a versioned executable inference interface or the complete final output transformation.
13 D / E	Class imbalance If class imbalance methods were used, state why and how this was done, and any subsequent methods to recalibrate the model or the model predictions	Methods 2.1 and 2.5; Table 1	[Partly reported] Class imbalance is documented and addressed in evaluation through macro-averaged metrics and class-stratified bootstrap sampling. DDCF training-time class weighting/resampling and subsequent recalibration are not specified. Class-weighted cross-entropy in the dataset report belongs to its own benchmark pipeline, not verified DDCF training.
14 D / E	Fairness Describe any approaches that were used to address model fairness and their rationale	Methods 2.5; Discussion — limitations Dataset [2]: Skin tone; Usage notes	[Not assessed] Demographic fairness was not assessed in the reported DDCF evaluation. The source dataset includes skin-tone metadata and reports underrepresentation of darker skin tones and children [2], but DDCF does not report sex-, age-, or skin-tone-specific performance or a demographic debiasing method.
15 D	Model output Specify the output of the prediction model (eg, probabilities, classification). Provide details and rationale for any classification and how the thresholds were identified	Methods 2.1 and 2.5; Results 3.5; Discussion	[Partly reported] The model predicts one of nine diagnostic categories. Evaluation distinguishes continuous class scores used for AUROC from final top-1 predictions used for BACC, macro-F1 and case transitions. A separately optimized clinical threshold, rejection rule or complete probability-transformation specification is not reported.
16 D / E	Training versus evaluation Identify any differences between the development and evaluation data in healthcare setting, eligibility criteria, outcome, and predictors	Methods 2.1; Table 1; Supplementary Table S1	[Partly reported] Development and evaluation use the same eligible source cohort, paired-image definitions and diagnostic categories. Stratification fixes class-wise allocation across repeated partitions; individual lesions differ. No distinct temporal/geographical evaluation cohort is used. Demographic and image-characteristic comparisons between training/validation/test subsets are not reported.

Ethics, open science and involvement

Item	Reporting requirement	Section / supporting source	Reporting information
17 D / E	Ethical approval Name the institutional research board or ethics committee that approved the study and describe the participant informed consent or the ethics committee waiver of informed consent	Ethics approval and consent to participate Dataset [2]: Ethics review	[Partly reported] DDCF reports secondary use of public de-identified data without new recruitment. Source collection approvals include Vienna 1081/2015, 1536/2017 and 1070/2023; Queensland 2015000162 and 2017001223; and MSK 16-974. Ankara/Skopje contributions used Vienna protocol 1536/2017. The source sought no written consent for nonidentifiable retrospective data [2]. These are not DDCF-specific ethics approvals; a separate local determination is not stated.
18a D / E	Funding Give the source of funding and the role of the funders for the present study	Funding	[Partly reported] The study reports National Research Foundation of Korea/MSIT support under 2022R1A2C2091160 and RS-2024-00440802. The funders' role in this DDCF study is not specified; the dataset article's funding statement does not apply to these authors' analysis.
18b D / E	Conflicts of interest Declare any conflicts of interest and financial disclosures for all authors	Competing interests	[Reported] The DDCF authors declare that they have no competing interests.
18c D / E	Protocol Indicate where the study protocol can be accessed or state that a protocol was not prepared	Protocol statement — not provided in current manuscript	[Not explicitly reported] The current DDCF manuscript does not state whether a separate study protocol was prepared or provide a protocol-access location. The source-data ethics protocols are not the DDCF model-development protocol.
18d D / E	Registration Provide registration information for the study, including register name and registration number, or state that the study was not registered	Clinical trial number	[Not explicitly reported] The manuscript states that a clinical trial number is not applicable, but it does not state whether this prediction-model study was registered in a broader study registry or explicitly state non-registration.
18e D / E	Data sharing Provide details of the availability of the study data	Availability of data and materials Dataset [2]: Data availability statement; Usage notes	[Reported] The manuscript cites the public MILK10k dataset. Its source report gives ISIC Archive DOI 10.34970/648456 and Harvard Dataverse DOI 10.7910/DVN/FSXRAQ; the dataset is distributed under CC BY-NC 4.0 for research/teaching [2]. DDCF-specific split-manifest availability is not specified.
18f D / E	Code sharing Provide details of the availability of the analytical code§	Availability of data and materials	[Reported] DDCF implementation code and analysis scripts are available from the corresponding author upon reasonable request. The source article's publicly linked baseline pipeline is a separate implementation and is not represented as the DDCF code release.
19 D / E	Patient and public involvement Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination of the study or state no involvement	Patient/public involvement statement — not provided in current manuscript	[Not explicitly reported] Patient and public involvement is not explicitly reported for the DDCF study. Retrospective use of a public dataset does not by itself establish whether patient/public involvement occurred in study design, interpretation, reporting, or dissemination.

Participants, model specification and results

Item	Reporting requirement	Section / supporting source	Reporting information
20a D / E	Participants Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful	Methods 2.1; Table 1 Dataset [2]: Table 1	[Reported] The lesion-level flow is 5,240 source lesions → removal of 44 other-benign and 9 other-malignant lesions → 5,187 eligible paired lesions → 3,630/519/1,038 training/validation/test lesions per partition. Diagnostic counts are provided. Lesions are the analysis units; a unique-patient count and longitudinal follow-up are not reported.
20b D / E	Participants Report the characteristics overall and, where applicable, for each data source or setting, including the key dates, key predictors (including demographics), treatments received, sample size, number of outcome events, follow-up time, and amount of missing data. A table may be helpful. Report any differences across key demographic groups	Methods 2.1; Table 1 Dataset [2]: Dataset description; Table 1	[Partly reported] For the full 5,240-lesion source cohort, the report gives mean age 61.1 years (SD 16.1), female records n = 2,088 (39.8%), six skin-tone levels, and anatomical-site counts, including 31 missing sites [2]. These describe the original source cohort, not recalculated demographics of the 5,187-lesion DDCF subset or its partitions. DDCF reports its own diagnostic-category counts.
20c E	Participants For model evaluation, show a comparison with the development data of the distribution of important predictors (demographics, predictors, and outcome)	Methods 2.1; Table 1; Supplementary Table S1	[Partly reported] The manuscript provides eligible-cohort and per-test-partition diagnostic counts plus training/validation/test totals. Fixed class-wise counts explain the common support across repeated partitions. Predictor/demographic distributions by development and test subset are not reported; source-cohort demographics cannot supply that comparison.
21 D / E	Model development Specify the number of participants and outcome events in each analysis (eg, for model development, hyperparameter tuning, model evaluation)	Methods 2.1 and 2.5; Table 1; Results 3.5; Supplementary Tables S1–S3	[Reported] Training, validation and test sizes are 3,630, 519 and 1,038 lesions per partition. Class-wise test support and the first-experiment 1,038-lesion case analysis are specified. Each paired bootstrap comparison uses its corresponding test partition. Separate class-wise training and validation counts are not tabulated.
22 D	Model specification Provide details of the full prediction model (eg, formula, code, object, application programming interface) to allow predictions in new individuals and to enable third party evaluation and implementation, including any restrictions to access or reuse (eg, freely available, proprietary) [¶]	Methods 2.2–2.4; Figure 1; Availability of data and materials	[Partly reported] The manuscript supplies the fusion equations, component architecture, backbone configurations and training/input-size settings. DDCF code is available on request. Trained-weight access, complete preprocessing and a versioned executable prediction interface are not specified; the dataset's separate baseline does not replace the DDCF model specification.
23a D / E	Model performance Report model performance estimates with confidence intervals, including for any key subgroups (eg, sociodemographic). Consider plots to aid presentation	Results 3.1–3.4; Tables 2–5; Supplementary Tables S1 and S3	[Partly reported] Model scores are reported as mean ± SD across partitions. DDCF–naïve-sum differences have partition-specific nominal 95% bootstrap intervals. These quantify paired differences, not absolute-model-performance or three-partition-mean CIs. Diagnostic-class results are reported; sociodemographic subgroup performance is not.
23b D / E	Model performance If examined, report results of any heterogeneity in model performance across clusters. See TRIPOD-Cluster for additional details [‡]	Methods 2.1 and 2.5; Discussion — limitations Dataset [2]: Usage notes	[Not assessed] Centre/country-specific heterogeneity of DDCF performance was not assessed. The source dataset withholds individual-image source labels [2]; repeated data partitions and diagnostic-category analyses are not cluster-level heterogeneity analyses.
24 E	Model updating Report the results from any model updating, including the updated model and subsequent performance	Results — fixed-configuration evaluation	[Not applicable] No post-evaluation updated or recalibrated model is reported; therefore there are no model-updating results to report for this evaluation.

Interpretation and applicability

Item	Reporting requirement	Section / supporting source	Reporting information
25 D / E	Interpretation Give an overall interpretation of the main results, including issues of fairness in the context of the objectives and previous studies	Discussion; Conclusions Dataset [2]: Usage notes	[Partly reported] Findings are interpreted in the context of asymmetric fusion, prior models, AUROC/BACC trade-offs, paired uncertainty and the BCC sensitivity–specificity trade-off. The source’s underrepresentation of darker skin tones and children limits demographic extrapolation [2]; no DDCF fairness conclusion follows from the aggregate comparisons.
26 D / E	Limitations Discuss any limitations of the study (such as a non-representative sample, sample size, overfitting, missing data) and their effects on any biases, statistical uncertainty, and generalisability	Discussion — limitations Dataset [2]: Usage notes; Data processing	[Reported] DDCF discusses conditional CIs, coefficient comparisons, unavailable patient identifiers, model-specific baselines, rare classes and a single retrospective dataset. The source further identifies verification/selection bias, limited darker-tone/pediatric coverage, anatomical underrepresentation and imperfect diagnostic ground truth [2]. These are source-data limitations, not newly measured DDCF effects.
27a D	Usability of the model in the context of current care Describe how poor quality or unavailable input data (eg, predictor values) should be assessed and handled when implementing the prediction model	Methods 2.1 and 2.2; Discussion — limitations Dataset [2]: Data collection and selection	[Not assessed] Deployment handling of missing, mismatched, or poor-quality paired inputs was not assessed. DDCF was evaluated on a quality-curated released dataset; source-level exclusion of poor-quality images [2] is not a tested DDCF deployment safeguard.
27b D	Usability of the model in the context of current care Specify whether users will be required to interact in the handling of the input data or use of the model, and what level of expertise is required of users	Background; Methods 2.1; Conclusions	[Not applicable] Not applicable to the current retrospective computational evaluation because no deployed clinical user interface or workflow was evaluated. Required operator expertise and user interaction would need to be specified in a future deployment study.
27c D / E	Usability of the model in the context of current care Discuss any next steps for future research, with a specific view to applicability and generalisability of the model	Discussion; Conclusions Dataset [2]: Usage notes	[Reported] The manuscript identifies independent, external and prospective evaluation as necessary before generalization or clinical utility can be established. Source-data limitations motivate testing across more representative demographic groups and anatomical sites [2]. Evaluation of other modality combinations and appropriate correction strengths is also discussed.

Requirement footnotes

† Applies separately to model-building approaches. ‡ See TRIPOD-Cluster for clustering/heterogeneity. § Analytical code includes data cleaning, feature engineering, model building and evaluation. ¶ Model specification enables predictions in a new individual.

References and source attribution

[1] Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi:10.1136/bmj-2023-078378.

[2] Tschandl P, Akay BN, Rosendahl C, et al. MILK10k: A Hierarchical Multimodal Imaging-Learning Toolkit for Diagnosing Pigmented and Nonpigmented Skin Cancer and its Simulators. *Journal of Investigative Dermatology*. 2026;146:357–364.e7. doi:10.1016/j.jid.2025.06.1594.

Reporting requirements follow Tables 2 and 3 of the TRIPOD+AI statement [1] (CC BY 4.0). Section locators and study-specific responses have been added. Source-data context is attributed separately to [2].