SCANSKINAI RESEARCH REPORT

Digital Skin Health Insights Report 2026

A descriptive scan-level analysis of ScanSkinAI risk bands and AI-supported medical skin-output patterns

Data coverage: 1 January 2026 – 28 July 2026

Version 1.0 | Published 29 July 2026

50,265 risk-band scan records • 45,270 medical skin-module scans

Scan-level descriptive analysis. Not a diagnostic-accuracy, prevalence or clinical-outcomes study.

Prepared by ScanSkinAI

Dr. Lifeng Zhu, PhD — Founder, ScanSkinAI

Publication information

Publication edition	2026
Data coverage	1 January 2026 – 28 July 2026
Data cut-off	28 July 2026
Data extraction date	29 July 2026
Publication date	29 July 2026
Version	1.0
Report type	Retrospective descriptive aggregate analysis
Unit of analysis	Scan records
Risk-band dataset	50,265 scan records
Medical skin-output dataset	45,270 medical skin-module scans
Report author	Dr. Lifeng Zhu, PhD — Founder, ScanSkinAI
Internal clinical review	ScanSkinAI Clinical Review Team — 29 July 2026
Internal technical/data review	ScanSkinAI Data & Engineering Team — 29 July 2026
Internal privacy/governance review	ScanSkinAI Privacy & Governance Function — 29 July 2026

> **Important interpretation notice**
>
> This report analyses scan records, not confirmed diagnoses or unique individuals. A person may have submitted more than one scan. AI-supported output labels and risk bands are preliminary screening information and must not be interpreted as clinical diagnoses, disease prevalence or diagnostic-accuracy results.

This report is provided for research transparency and general educational purposes. ScanSkinAI outputs are preliminary screening information and are not medical diagnoses. A low-risk or normal-skin output cannot rule out disease. Anyone concerned about a new, changing, painful, itching, bleeding, crusting or non-healing skin mark should seek advice from an appropriately qualified healthcare professional.

Contents

Publication information 1
Executive summary 3
1. Background and purpose 4
2. Dataset overview 4
3. Results 5
4. Interpretation and implications 8
5. Privacy and responsible aggregate reporting 9
6. Limitations 9
7. Conclusions 10
Part II. Methodology 11
8. Methodology 12
9. Research governance, funding and integrity 15
10. References 16
Appendix A. Complete aggregate data tables 17
Appendix B. Glossary 18

Executive summary

This report presents a retrospective descriptive analysis of approved aggregate ScanSkinAI scan-history data generated through routine platform use between 1 January and 28 July 2026. The unit of analysis is a scan record, not a person. It is therefore essential to read every result as a distribution of platform outputs rather than a count of unique users or clinically confirmed conditions.

Two separate analysis datasets are described. The risk-band distribution included 50,265 scan records. The AI-output distribution included 45,270 medical skin-module scans. These datasets have different denominators and must not be combined. The AI-output distribution was available for a narrower eligible subset of medical skin-module scans; no further record-level breakdown is published.

Within the medical skin-module subset, the most frequent individually named AI-supported output category was benign naevus / mole, representing 13.9% of module scans. “Normal skin” represented 11.8%, followed by contact dermatitis at 7.3%, basal cell carcinoma indication at 6.5%, acne vulgaris at 5.9% and melanoma indication at 5.0%. The cancer-related categories retain the word “indication” because they are preliminary AI-supported outputs, not diagnoses.

Approximately 26% of medical skin-module scans were distributed across approximately 430 other lower-frequency labels. This broad long tail shows that the platform output distribution extends well beyond the most frequent named categories; it does not establish the prevalence of approximately 430 diseases.

Within the risk-band dataset, 52.8% of scan records fell in the 1–39 low band, 27.7% in the 40–69 moderate band and 18.6% received a 70–100 high-concern flag. A small proportion recorded a score of zero (0.7%) or had no reportable risk-band value (0.3%). A high-concern flag does not confirm cancer, while a low score does not rule out a medical condition.

The findings describe platform activity only. They cannot be used to estimate disease prevalence or incidence, evaluate diagnostic accuracy, count confirmed cancers, establish clinical outcomes, or describe risk in the general population. No individual user records, identifiers, images or row-level data are published.

50,265 Risk-band scan records	**45,270** Medical skin-module scans
52.8% Scan records in the 1–39 low band	**≈26%** Medical outputs across ≈430 other labels

> **Bottom line**
>
> The report is valuable as a transparent account of how ScanSkinAI’s preliminary output categories and risk bands were distributed across an approved aggregate scan extract. Its value depends on keeping the claims strictly scan-level, non-diagnostic and descriptive.

1. Background and purpose

Digital skin-health services generate large volumes of operational data through routine use. Aggregate reporting can improve product transparency, support responsible communication and identify questions for future research. However, routine platform data are not automatically equivalent to clinical research data: the population is self-selected, outputs may not be clinically confirmed, users may scan repeatedly, and product changes can affect the observed distribution.

The purpose of this report is narrow and descriptive. It summarises two approved aggregate extracts: a risk-band distribution covering 50,265 scan records and an AI-output distribution covering 45,270 medical skin-module scans. It does not evaluate ScanSkinAI's diagnostic performance; that question belongs to separate validation work.

1.1 Research objectives

1. Describe the distribution of available risk bands across approved scan records.
2. Describe the most frequent preliminary AI-supported output categories within the medical skin-module subset.
3. Describe the scale of lower-frequency output labels grouped into the long-tail category.
4. Explain the limits of interpreting scan-level AI outputs as health or population statistics.
5. Provide transparent information about data scope, privacy and methodological constraints.

No hypotheses were specified and no inferential statistical analyses were performed.

1.2 Scope boundaries

The report intentionally excludes demographic, geographic, seasonal and outcome analyses because those variables were not approved for public disclosure within the aggregate extract. It also excludes unique-user counts, model-version-level comparisons, sensitivity, specificity, predictive values, histopathology confirmation and clinical follow-up.

2. Dataset overview

2.1 Two datasets, two denominators

Risk-band analysis	50,265 scan records; denominator for every risk-band percentage.
Medical skin-module AI-output analysis	45,270 scans; denominator for every AI-output percentage.
Relationship	The AI-output table represents a narrower eligible medical skin-module subset. No further record-level breakdown is published.
Unit	Scan records. The number of unique people is not established.

Do not combine the denominators

A percentage calculated from the 50,265-record risk-band dataset must never be expressed as a share of the 45,270-scan AI-output dataset, or vice versa.

2.2 Definitions used throughout the report

Term	Definition
Scan record	One eligible processed scan record represented in the approved aggregate extract.
Risk-band record	A scan record with a risk score or missing-risk status included in the 50,265-record dataset.
Medical skin-module output record	A scan included in the 45,270-record medical skin AI-output distribution.

Term	Definition
AI-supported output category	A preliminary model-generated label; not a biopsy-, pathology- or clinician-confirmed diagnosis in this analysis.
Unique user	Not available in the approved public aggregate dataset.
Confirmed diagnosis	Not assessed in this report.

3. Results

3.1 Risk-band distribution

Risk bands are platform output categories applied to 50,265 scan records. They are not a clinical staging system, a cancer diagnosis or a validated population-risk classification.

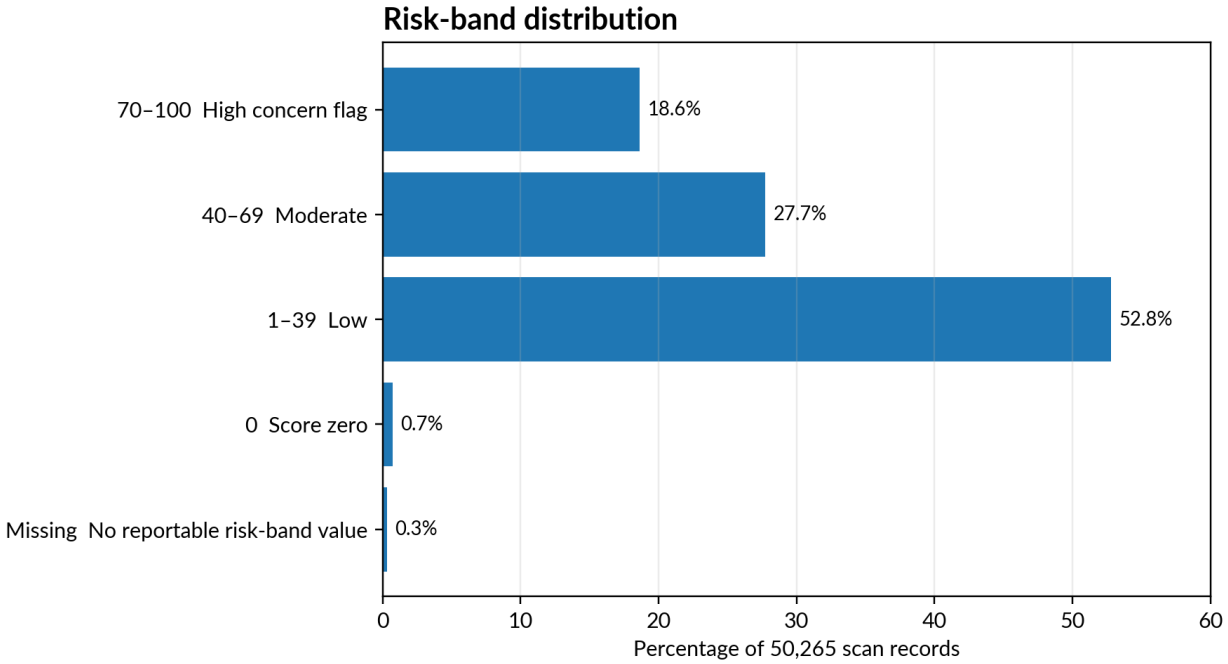

Figure 1. Risk-band distribution across 50,265 scan records. Percentages are rounded.

Risk band	Category	Scan records	% of 50,265
70–100	High concern flag	9,323	18.6%
40–69	Moderate	13,904	27.7%
1–39	Low	26,538	52.8%
0	Score zero	353	0.7%
Missing	No reportable risk-band value	147	0.3%

Interpretation

A high-concern flag is a screening output, not a confirmed cancer diagnosis. A low score does not rule out disease. These figures are not cancer rates and are not measures of disease prevalence.

3.2 Medical skin-module AI-output distribution

The AI-output analysis included 45,270 medical skin-module scans. The labels represent preliminary model-generated categories and not clinician-confirmed diagnoses. Cancer-related categories are therefore reported as “indications”.

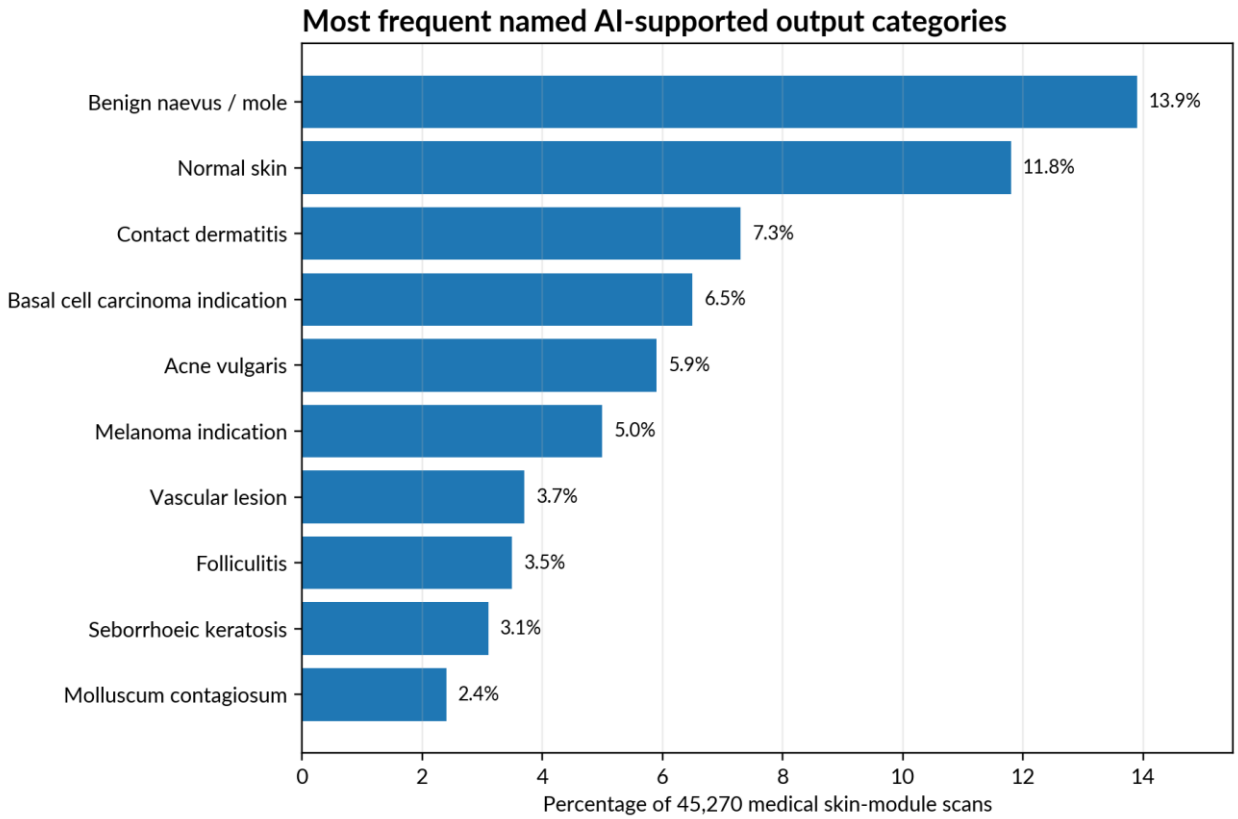

Figure 2. Ten most frequent named AI-supported output categories. Denominator: 45,270 medical skin-module scans.

Benign naevus / mole was the most frequent individually named category at 13.9%, followed by normal skin at 11.8% and contact dermatitis at 7.3%. Basal cell carcinoma indication represented 6.5%, acne vulgaris 5.9% and melanoma indication 5.0%. These proportions describe AI-output categories among scan records; they do not describe the proportion of users with those conditions.

3.2.1 Complete approved AI-output table

AI-supported output group	Scan count	% of module	Interpretation
Benign naevus / mole	6,270	13.9%	Preliminary AI-supported output; not a clinical confirmation.
Normal skin	5,336	11.8%	Does not guarantee the absence of disease.
Contact dermatitis	3,309	7.3%	Preliminary AI-supported output; not a clinical confirmation.
Basal cell carcinoma indication	2,932	6.5%	Concern indication only; not confirmed basal cell carcinoma.
Acne vulgaris	2,688	5.9%	Preliminary AI-supported output; not a clinical confirmation.
Melanoma indication	2,273	5.0%	Concern indication only; not confirmed melanoma.
Vascular lesion	1,683	3.7%	Preliminary AI-supported output; not a clinical confirmation.

AI-supported output group	Scan count	% of module	Interpretation
Folliculitis	1,593	3.5%	Preliminary AI-supported output; not a clinical confirmation.
Seborrhoeic keratosis	1,402	3.1%	Preliminary AI-supported output; not a clinical confirmation.
Molluscum contagiosum	1,081	2.4%	Preliminary AI-supported output; not a clinical confirmation.
Dermatofibroma	984	2.2%	Preliminary AI-supported output; not a clinical confirmation.
Eczema / atopic dermatitis	835	1.8%	Preliminary AI-supported output; not a clinical confirmation.
Tinea corporis	825	1.8%	Preliminary AI-supported output; not a clinical confirmation.
Impetigo	796	1.8%	Preliminary AI-supported output; not a clinical confirmation.
Squamous cell carcinoma indication	Not publicly reported	1.2%	Concern indication only; not confirmed squamous cell carcinoma.
Actinic keratosis	516	1.1%	Preliminary AI-supported output; not a clinical confirmation.
Not a skin image / unclassified	305	0.7%	Image could not be assigned to a reportable skin-output category.
All other labels	≈11,900	≈26%	Approximately 430 lower-frequency output labels grouped for readable reporting.

The exact approved count for squamous cell carcinoma indication was not included in the supplied aggregate extract. The report therefore shows the approved percentage only and does not reverse-calculate a count from a rounded percentage.

3.3 A broad long tail of output categories

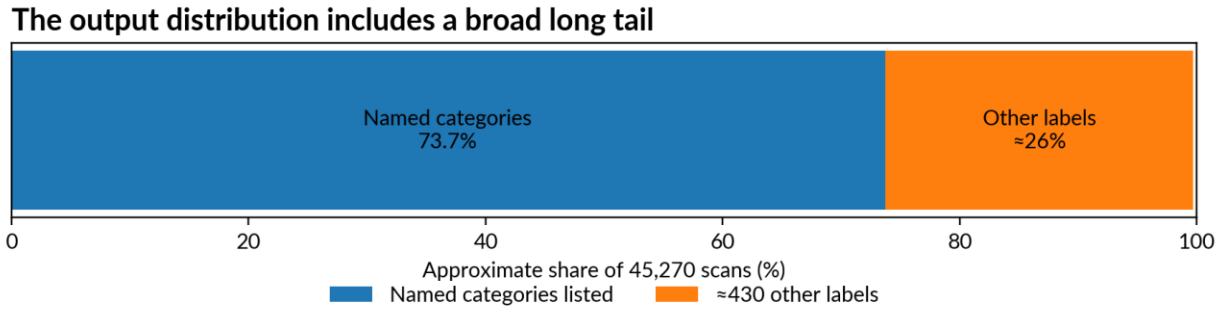

Figure 3. Approximately 26% of medical skin-module scans were distributed across approximately 430 other labels. Values are approximate and rounded.

Approximately 11,900 scans—approximately 26% of the 45,270-scan medical skin-module dataset—were distributed across approximately 430 other labels. Grouping these lower-frequency outputs improves readability, but reduces category-level detail. The grouped result is not one clinical category and does not establish the prevalence of approximately 430 diseases.

The long tail may reflect the breadth of the model taxonomy, differing user concerns, repeated scans and the mix of platform use cases. It should be treated as a descriptive feature of the output distribution rather than a measure of clinical burden.

4. Interpretation and implications

4.1 What the findings may indicate

- Benign naevus / mole was the most frequent individually named AI-output category in the approved extract.
- The output distribution includes both common inflammatory concerns and lesion-related indication categories.
- A substantial proportion of records is distributed across a long tail of lower-frequency labels.
- More than half of risk-band scan records were in the low band.
- Moderate and high-concern bands require responsible next-step communication because they remain screening outputs rather than diagnoses.
- The aggregate distribution can inform product transparency, user education, quality review and future research planning.

4.2 What the findings cannot show

- The number of unique people represented.
- Disease prevalence or incidence.
- Confirmed diagnoses or confirmed cancer cases.
- Diagnostic accuracy, sensitivity, specificity or predictive values.
- Clinical outcomes or whether users followed medical advice.
- Whether repeat scans came from the same skin concern.
- Age, sex, gender, skin-tone, geographic or seasonal differences.
- Changes over time or causes of the observed distribution.
- Risk in the general population.

> **Critical distinction**
>
> This report describes what the platform output, not what a clinician ultimately diagnosed. The correct language is “AI-supported output category” or “indication”, never “confirmed case”.

4.3 Responsible communication of cancer-related indications

The categories “basal cell carcinoma indication”, “melanoma indication” and “squamous cell carcinoma indication” must retain the word “indication” in every chart, table, press release and social-media post based on this report. The categories must not be added together to claim a number of cancers detected, because the analysis contains no linked clinical confirmation and may include repeated scans.

Similarly, the “normal skin” output must not be communicated as evidence that no condition is present. A person with a changing or symptomatic mark should seek appropriate professional advice regardless of a low-risk or normal-skin output.

5. Privacy and responsible aggregate reporting

The public report is intentionally limited to approved aggregate statistics. It does not publish raw photographs, facial images, names, contact information, IP addresses, individual scan histories, free-text notes, pseudonymous identifiers or row-level data. The displayed tables are designed not to allow readers to trace a statistic to an identifiable person.

The reporting approach is consistent with the principle that research and statistical processing should use anonymous information where possible and otherwise apply data minimisation and appropriate safeguards. Pseudonymised information remains personal data where re-identification is possible using additional information; this report therefore avoids releasing pseudonymous row-level data.[3–5]

Area	Publication approach
Published	Aggregate counts, percentages, definitions, limitations, methodology and version history.
Not published	Identifiers, images, row-level records, user histories, free text, partner-specific or employer-specific breakdowns.
Individual decisions	The report is not used to make employment, insurance or healthcare decisions about named people.

Area	Publication approach
Underlying data	Production data remains restricted and is not released as part of this publication.

This publication does not itself constitute a legal-compliance determination. Appropriate lawful-basis, special-category-data and governance assessments remain organisational responsibilities and should be documented separately.

6. Limitations

1. Scan-level denominator. A person may contribute more than one scan, so the counts describe scans and not people.

2. Unknown unique-user count. The approved aggregate data does not establish how many unique people are represented; no per-person rate can be calculated.

3. Self-selected platform population. People using ScanSkinAI are not a random sample of the general population and may scan because a mark is already concerning.

4. AI-output status. Output labels are preliminary model-generated categories and may be wrong in either direction.

5. No clinical ground-truth analysis. The report does not compare outputs with biopsy, pathology or dermatologist-confirmed diagnoses.

6. No accuracy evaluation. The report cannot be used to infer sensitivity, specificity, predictive value, diagnostic accuracy or clinical utility.

7. Different analysis denominators. Risk-band and AI-output results came from different eligible subsets, so cross-table arithmetic would be invalid.

8. Repeated use. Repeat scans may influence category counts, and the report cannot determine whether repeated scans represent the same concern.

9. Limited public variables. Age, geography, skin type, device, sex, symptoms and follow-up outcomes are not included.

10. Model and taxonomy changes. If more than one product or model version is represented, output labels and thresholds may have shifted during the period.

11. Eligibility and deduplication verification. The aggregate extract did not support independent verification of every record-level eligibility, test-account or retry-exclusion rule.

12. Approximate values and rounding. The long-tail count and share are approximate, and displayed percentages may not sum to exactly 100%.

13. Long-tail grouping. Grouping approximately 430 labels reduces detail and may obscure differences between low-frequency categories.

14. No population inference. Results cannot be converted into national or global prevalence or incidence estimates.

15. No causal conclusions. The report describes distributions only and cannot explain why the distribution occurred.

16. Company-authored analysis. ScanSkinAI developed the technology and prepared the report, so the publication should not be interpreted as an independent external evaluation.

7. Conclusions

Across the approved 50,265-record risk-band extract, the low band accounted for the largest share of scan records, while moderate and high-concern outputs together represented a substantial minority. Across 45,270 medical skin-module scans, benign naevus / mole was the most frequent individually named AI-output category, followed by normal skin and contact dermatitis. Approximately one quarter of module scans fell across a broad long tail of approximately 430 other labels.

These findings provide a transparent description of ScanSkinAI platform outputs during the stated data period. They do not establish diagnoses, disease rates, accuracy or clinical outcomes. Their responsible use is therefore limited to product transparency, education and the generation of carefully designed future research questions.

7.1 Recommended next research steps

- Establish a reproducible record-level eligibility and deduplication pipeline for future annual reports.
- Report unique-user counts separately from scan counts when governance permits.
- Document model and taxonomy versions within each reporting period.
- Pre-register future descriptive questions and publication rules before inspecting results.

- Undertake separately designed clinical validation studies with appropriate reference standards when evaluating accuracy.
- Consider independent statistical, clinical and privacy review for future editions.

PART II

Research Methodology

Data scope, scan-level definitions, aggregate analysis methods, privacy safeguards and interpretation limits

8. Methodology

8.1 Methodology summary

This project was a retrospective descriptive analysis of approved aggregate ScanSkinAI scan-history data. It describes distributions of scan-level risk bands and preliminary medical skin-module AI-output categories. It was not designed to evaluate diagnostic accuracy, determine disease prevalence, establish incidence, measure clinical outcomes or draw causal conclusions.

The public reporting structure was informed by STROBE recommendations for transparent reporting of observational studies and the RECORD extension for studies using routinely collected health data.[1,2] These frameworks guide reporting completeness; they do not validate the underlying data, study design or conclusions.

8.2 Study design and setting

Study design: retrospective descriptive aggregate analysis of routinely collected platform scan records. Data was generated through normal ScanSkinAI platform use rather than through a prospective research protocol. No users were recruited for this report, no intervention was assigned, and there was no control group or follow-up arm.

Platform	ScanSkinAI
Data coverage	1 January 2026 – 28 July 2026
Data cut-off	28 July 2026
Data extraction date	29 July 2026
Analysis type	Descriptive aggregate analysis
Public data level	Approved aggregate counts and percentages only
Model-version reporting	Not reported publicly

8.3 Analysis populations

Population	Denominator	Purpose
Population A: risk-band analysis	50,265 scan records	Used only for risk-band counts and percentages.
Population B: medical skin-module outputs	45,270 scans	Used only for AI-output counts and percentages.

The AI-output distribution was available for a narrower eligible subset of medical skin-module scans. The public aggregate extract does not provide a further record-level breakdown of the 4,995-record difference. Results from the two populations are never combined, and no figure from one table is expressed as a share of the other.

8.4 Inclusion and exclusion information

Included records were those contained in the approved aggregate production extract, within the verified data-coverage period, and represented in the relevant risk-band or medical skin-module output table. Records outside the approved reporting scope, outside the medical module for the output table, or unavailable in the approved aggregate extract were not represented in the relevant analysis.

The public aggregate data does not permit independent verification of every record-level eligibility and deduplication rule. No affirmative claim is therefore made that every staff, demo, bot, retry or sandbox record was excluded. This constraint is reported as a limitation rather than resolved through assumption.

8.5 Risk-band definitions

Band	Definition
70–100	High concern flag
40–69	Moderate
1–39	Low
0	Score zero
Missing	No reportable risk-band value

Risk bands are platform output categories. They are not clinical staging systems, cancer diagnoses, or validated population-risk categories. This report does not evaluate the performance of the thresholds or compare them with clinical disease severity.

8.6 AI-output definitions

- AI-output groups are preliminary model-generated labels and are not necessarily user-selected concerns.
- No biopsy, pathology or dermatologist confirmation was linked to the records within this analysis.
- A label ending in “indication” signals a preliminary concern output, never a confirmed diagnosis.
- A “normal skin” output does not rule out disease.
- “Not a skin image / unclassified” represents an image without a reportable skin category.
- Approximately 430 lower-frequency labels were grouped into “all other labels”.

8.7 Statistical methods

Only descriptive statistics were used: counts, percentages, ordered category tables, horizontal bar charts, a stacked risk-band chart and approximate grouped totals where explicitly marked. Percentages were calculated as the number of scan records in a category divided by the relevant analysis denominator, multiplied by 100.

Item	Method
Risk-band percentages	Category scan count ÷ 50,265 × 100
Medical skin-output percentages	Category scan count ÷ 45,270 × 100
Displayed precision	One decimal place
Approximate source values	Retained with the approximation symbol (≈)
Unreported SCC count	Not reverse-calculated from the rounded 1.2% value

No hypothesis tests, p-values, confidence intervals, regression, causal modelling, accuracy calculations, sensitivity or specificity calculations, population weighting, imputation, age standardisation or geographic adjustment were performed.

8.8 Missing, unclassified and approximate values

The risk-band dataset contained 147 missing values (0.3%), meaning there was no reportable risk-band value in the approved extract. The medical skin-module table contained 305 “not a skin image / unclassified” records (0.7%), meaning the image could not be assigned to a reportable skin-output category. These are different concepts and were not combined.

Approximately 430 lower-frequency labels were grouped into “all other labels”, representing approximately 11,900 scans and approximately 26% of the medical skin-module dataset. The figures are reproduced as approximate and are not treated as exact.

8.9 Privacy and disclosure control

The publication is limited to approved aggregate statistics. Raw records, images, identifiers and row-level data are not made publicly available to protect privacy, security, contractual confidentiality and proprietary system information. This approach reflects data-minimisation principles and the preference for anonymous information where the research purpose can be achieved without identifiable data.[3–5]

- No direct identifiers
- No photographs or facial images
- No free-text notes
- No individual medical histories
- No user-level export
- No pseudonymous identifiers
- No small demographic or geographic subgroup tables
- No employer- or partner-specific breakdowns
- No row-level public data

8.10 Bias and interpretive constraints

Source of bias or constraint	Interpretive effect
Selection and self-selection	The distribution reflects people and organisations choosing to use ScanSkinAI, often because a skin concern already exists.
Repeated records	One person may contribute several scans, potentially over-representing persistent concerns or frequent users.
Misclassification	Preliminary AI labels can be incorrect in either direction.
Model and taxonomy changes	If multiple versions are represented, changes to labels or thresholds may affect comparability.
Unmeasured external factors	Marketing activity, partner launches and interface changes may influence scan volume and category mix.
No external outcomes	Care-seeking and clinical conclusions outside the platform are not observable in this dataset.

8.11 Data availability and reproducibility

The underlying row-level dataset is not publicly available because it may contain sensitive health-related platform information, images, contractual information and proprietary system data. Publicly available materials comprise the aggregate tables in this report, the methodology, definitions, version history and correction process.

A future aggregate-data file may be published only if it contains exactly the approved public tables, has passed privacy and governance review, and contains no user-level, scan-level, pseudonymous, exact-timestamp, image, note or partner-specific information.

8.12 Reporting-standards statement

This report was structured with reference to relevant STROBE and RECORD transparency principles for observational analyses and routinely collected health data.[1,2] These frameworks guide transparent reporting but do not themselves validate the research design, data quality or conclusions. No “STROBE certification” or “RECORD approval” is claimed.

9. Research governance, funding and integrity

9.1 Authorship and internal review

Role	Details

Role	Details
Report author and data-analysis owner	Dr. Lifeng Zhu, PhD — Founder, ScanSkinAI
Internal clinical review	ScanSkinAI Clinical Review Team — reviewed 29 July 2026
Internal technical/data review	ScanSkinAI Data & Engineering Team — reviewed 29 July 2026
Internal privacy/governance review	ScanSkinAI Privacy & Governance Function — reviewed 29 July 2026

Clinical review covers the wording of the findings and safety framing only; it does not verify every underlying record or confirm any diagnosis. Unless specifically stated, the report should not be interpreted as an independent external evaluation.

9.2 Funding and commercial interest

This report was prepared and funded by ScanSkinAI. ScanSkinAI developed and operates the technology from which the aggregate statistics were generated. Commercial involvement and the limitations of company-controlled platform data should be considered when interpreting the findings.

9.3 Ethics and peer review

No ethics-committee approval and no external academic peer review are claimed for this report. It is a company-authored descriptive transparency report based on approved aggregate platform data.

9.4 Corrections and versioning

Substantive corrections will increment the version number, be described in the version history and update the modification date. Material results will not be silently replaced. Correction requests may be sent to info@scanskinai.com.

Version	Date	Change
1.0	29 July 2026	Initial publication

9.5 How to cite this report

> **Suggested citation**
>
> Zhu L. ScanSkinAI Digital Skin Health Insights Report 2026: A Descriptive Scan-Level Analysis of Risk Bands and AI-Supported Medical Skin Outputs. Version 1.0. Published 29 July 2026. ScanSkinAI.

Report URL: https://www.scanskinai.com/research/digital-skin-health-report-2026

Methodology URL: https://www.scanskinai.com/research/digital-skin-health-report-2026-methodology

No DOI has been assigned to this Digital Skin Health Insights Report. The Zenodo identifiers listed in the references belong to the separate ScanSkinAI Cancer Flag Module technical whitepaper. Use DOI 10.5281/zenodo.21225286 to cite all versions of that whitepaper, or DOI 10.5281/zenodo.21225287 to cite the specific archived version referenced here.

10. References

1. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. PLoS Medicine. 2007;4(10):e296. doi:10.1371/journal.pmed.0040296.

https://doi.org/10.1371/journal.pmed.0040296

2. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al.; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Medicine. 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885.

https://doi.org/10.1371/journal.pmed.1001885

3. Information Commissioner's Office. The research provisions: What are the appropriate safeguards? Accessed 29 July 2026.

https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/the-research-provisions/what-are-the-appropriate-safeguards/

4. Information Commissioner's Office. Pseudonymisation. Accessed 29 July 2026.

https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/pseudonymisation/

5. Information Commissioner's Office. Principles and grounds for processing research-related data. Accessed 29 July 2026.

https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/the-research-provisions/principles-and-grounds-for-processing/

10.1 ScanSkinAI research and technical sources

6. Ivy AI Solutions Limited. ScanSkinAI Cancer Flag Module: Training Methodology, Model Development, and Robustness Evaluation. Zenodo; 2026. Use the concept DOI to cite all versions; the version DOI identifies the specific archived release referenced in this report.

All versions (concept DOI): https://doi.org/10.5281/zenodo.21225286

Specific archived version: https://doi.org/10.5281/zenodo.21225287

Zenodo record: https://zenodo.org/records/21225287

7. ScanSkinAI. ScanSkinAI Accuracy Report 2026: Independent Clinical Validation of AI Skin Screening. Ivy AI Solutions Limited; 2026. Accessed 29 July 2026.

https://www.scanskinai.com/accuracy-report

8. ScanSkinAI. AI Skin Cancer Screening Methodology (Whitepaper Reference Page). Ivy AI Solutions Limited; 2026. Accessed 29 July 2026.

https://www.scanskinai.com/research/ai-skin-cancer-screening-methodology

9. ScanSkinAI. The Cancer Flag Module: Non-Diagnostic AI Triage for Skin Cancer from a Photo. Ivy AI Solutions Limited; 2026. Accessed 29 July 2026.

https://www.scanskinai.com/research/cancer-flag-module

10. ScanSkinAI. Skin Cancer Statistics 2026: UK, US & Australia. Ivy AI Solutions Limited; 2026. Accessed 29 July 2026.

https://www.scanskinai.com/skin-cancer-statistics

11. ScanSkinAI. AI Skin Cancer Detection - Evidence & Clinical Validation. Ivy AI Solutions Limited; 2026. Accessed 29 July 2026.

https://www.scanskinai.com/evidence-base

12. ScanSkinAI. Research & Evidence Hub. Ivy AI Solutions Limited; 2026. Accessed 29 July 2026.

https://www.scanskinai.com/research

Zenodo citation note: Zenodo assigns a DOI to each published version and a concept DOI to the complete version chain. The concept DOI above is therefore the preferred identifier when the intention is to cite all versions; the version DOI should be used when reproducibility requires the exact archived files reviewed.

Appendix A. Complete aggregate data tables

A.1 Risk-band data

Risk band	Category	Scan records	Percentage
70–100	High concern flag	9,323	18.6%
40–69	Moderate	13,904	27.7%
1–39	Low	26,538	52.8%

Risk band	Category	Scan records	Percentage
0	Score zero	353	0.7%
Missing	No reportable risk-band value	147	0.3%

Denominator: 50,265 scan records. Percentages are rounded. Missing values remain in the denominator and are shown separately.

A.2 Medical skin-module AI-output data

AI-supported output group	Scan count	Percentage	Interpretation
Benign naevus / mole	6,270	13.9%	Preliminary AI-supported output; not a clinical confirmation.
Normal skin	5,336	11.8%	Does not guarantee the absence of disease.
Contact dermatitis	3,309	7.3%	Preliminary AI-supported output; not a clinical confirmation.
Basal cell carcinoma indication	2,932	6.5%	Concern indication only; not confirmed basal cell carcinoma.
Acne vulgaris	2,688	5.9%	Preliminary AI-supported output; not a clinical confirmation.
Melanoma indication	2,273	5.0%	Concern indication only; not confirmed melanoma.
Vascular lesion	1,683	3.7%	Preliminary AI-supported output; not a clinical confirmation.
Folliculitis	1,593	3.5%	Preliminary AI-supported output; not a clinical confirmation.
Seborrhoeic keratosis	1,402	3.1%	Preliminary AI-supported output; not a clinical confirmation.
Molluscum contagiosum	1,081	2.4%	Preliminary AI-supported output; not a clinical confirmation.
Dermatofibroma	984	2.2%	Preliminary AI-supported output; not a clinical confirmation.
Eczema / atopic dermatitis	835	1.8%	Preliminary AI-supported output; not a clinical confirmation.
Tinea corporis	825	1.8%	Preliminary AI-supported output; not a clinical confirmation.
Impetigo	796	1.8%	Preliminary AI-supported output; not a clinical confirmation.
Squamous cell carcinoma indication	Not publicly reported	1.2%	Concern indication only; not confirmed squamous cell carcinoma.
Actinic keratosis	516	1.1%	Preliminary AI-supported output; not a clinical confirmation.
Not a skin image / unclassified	305	0.7%	Image could not be assigned to a reportable skin-output category.

AI-supported output group	Scan count	Percentage	Interpretation
All other labels	≈11,900	≈26%	Approximately 430 lower-frequency output labels grouped for readable reporting.

Denominator: 45,270 medical skin-module scans. Approximate values are marked ≈. The exact scan count for squamous cell carcinoma indication is not publicly reported and has not been estimated from the rounded percentage.

Appendix B. Glossary

Term	Meaning in this report
Aggregate data	Information summarised across groups of records rather than displayed at individual-record level.
AI-supported output category	A preliminary model-generated label produced by the platform; not a diagnosis in this report.
Concern indication	A preliminary output category that may support escalation or professional review; not a confirmed malignancy.
Descriptive analysis	An analysis that summarises observed counts or distributions without testing causal hypotheses.
Incidence	The occurrence of new cases in a defined population over a defined period; not measured in this report.
Prevalence	The proportion of a defined population with a condition; not measured in this report.
Pseudonymised data	Personal data processed so it cannot be attributed to a person without additional information; it remains personal data where re-identification is possible.
Risk band	A ScanSkinAI platform output range used for preliminary concern communication; not clinical staging.
Scan record	One eligible processed scan represented in the aggregate extract.
Unique user	A distinct person or account. The number is not available in the approved public aggregate data.

Medical and interpretation disclaimer

This report is provided for research transparency and general educational purposes. ScanSkinAI outputs are preliminary screening information and are not medical diagnoses. A low-risk or normal-skin output cannot rule out disease. Anyone concerned about a new, changing, painful, itching, bleeding, crusting or non-healing skin mark should seek advice from an appropriately qualified healthcare professional.