Skin2Mind: A Multimodal Framework for Mental Health Risk Screening via Facial Images and Structured Skin Reports

Jiahe Liu[1], Duong Nhu[1], Xiangyu Zhao[1], Yiwen Jiang[1], Siyuan Yan[1], Xiangyi Xue[1], Yaling Shen[1], Guilherme Camargo de Oliveira[1], Stephanie Fong[1], Qingyang Xu[1], Zimu Wang[1], Fudi Wang[3,4], Xiaoqing Zhang[5], Dominic Dwyer[1,2], Levin Kuhlmann[1], and Zongyuan Ge[1]

[1] Monash University, Melbourne, Australia
[2] The University of Melbourne, Melbourne, Australia
[3] CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai, China
[4] EveLab Insight (Singapore) Pte. Ltd., Singapore
[5] West China Fourth Hospital, Sichuan University, Chengdu, China
{grace.liu; zongyuan.ge}@monash.edu

Abstract. Mental health screening predominantly relies on subjective questionnaires, leaving observable biological signals uncaptured. The skin-brain axis, mediated by neuro-immune and endocrine pathways, links psychological states to measurable changes in facial skin phenotypes. However, no computational framework has directly modeled this link for individual-level mental health risk prediction. We propose Skin2Mind, to the best of our knowledge, the first psychodermatology-grounded framework that integrates multi-spectral facial skin images (white, cross-polarized, and UV illumination) with structured skin analysis reports covering pigmentation, pilosebaceous activity, vascular response, structural aging patterns, barrier sensitivity, and facial morphology. To address inter-modal redundancy in co-acquired skin imaging and structured report modalities, we introduce a Hierarchical Anchor Residual Fusion strategy that treats UV images as the anchor modality and integrates auxiliary modalities through gated residual pathways. Evaluated on a clinical cohort of 192 subjects, Skin2Mind achieves an Area Under the ROC Curve (AUC) of 81.66% for DSM-5-aligned at-risk mental health classification, outperforming representative tabular, unimodal, and vision-language baselines. More importantly, semantic-level Leave-One-Feature-Out analysis reveals that the model predictions are driven by dermatologic dimensions consistent with skin-brain axis mechanisms, including epidermal barrier sensitivity, periorbital fatigue cues, pilosebaceous activity, and stress-related structural aging features. These findings provide the first computational evidence that quantitative facial skin phenotypes can serve as non-invasive, scalable biomarkers for mental health risk screening.

Keywords: Mental Health Screening · Multimodal Learning · Skin Phenotype

1 Introduction

Mental health risk screening predominantly relies on self-reported questionnaires and clinical interviews [7]. While these remain the clinical gold standard, they depend on patient disclosure and clinician interpretation, which can be affected by recall bias, social desirability, and limited patient insight [15]. As demand for early mental health intervention grows, there is increasing interest in complementary digital biomarkers that capture physiological correlates beyond patient self-report [16]. The skin presents a unique and accessible opportunity for such screening, acting as a visible interface that reflects internal psychological states through external physiological changes [14]. This connection is formalized through the skin-brain axis, which is biologically grounded in the neuro-immuno-cutaneous system (NICS) [21]. The NICS conceptualizes the skin as an active sensory organ linking neural, immune, and endocrine networks [12, 14], suggesting that emotional distress and mental health conditions can manifest as measurable facial skin phenotypes.

Despite this biological plausibility, facial skin phenotypes remain underexplored for mental health screening. Recent AI for dermatology studies have begun to utilize quantitative skin phenotypes to analyze visible skin symptoms in cohorts exposed to prolonged stress [4]. However, such epidemiological research treats psychological stress as an environmental assumption rather than utilizing paired, individual-level ground-truth labels. Without standardized psychiatric assessments (e.g., DSM-5-based questionnaires), existing work cannot directly link quantitative skin phenotypes to mental health risk at the individual-level. This reveals a critical research gap: the lack of computational frameworks that directly model the relationship between quantitative facial skin phenotypes and individual-level mental health risk.

To bridge this gap, we present Skin2Mind, the first end-to-end multimodal framework that directly uses skin phenotypes to predict individual-level, DSM-aligned mental health risk. Skin2Mind integrates multi-spectral facial skin images (cross-polarized, UV, and white-light) with structured skin reports automatically derived from the same imaging system. As the multispectral image modalities and structured skin reports are derived from the same acquisition pipeline, they share partially overlapping information. To integrate them effectively, it is necessary to learn discriminative representations for each modality and adopt fusion strategies that mitigate redundancy. Thus, we employ a domain-specific dermatology foundation model [22] to encode images and a Large Language Model (LLM) [17] to encode text reports. To effectively integrate these embeddings, we propose a novel Hierarchical Anchor Residual Fusion (HARF) strategy. HARF employs a two-stage anchoring design: it first anchors on the most stable UV image to inject auxiliary image cues through gated residual pathways, and subsequently treats the unified image representation as the anchor for integrating semantic text features. To enhance clinical interpretability, we perform a semantic-level Leave-One-Feature-Out (LOFO) analysis, demonstrating that learned decision patterns align with established skin-brain axis phenotypes.

In summary, our main contributions are threefold:

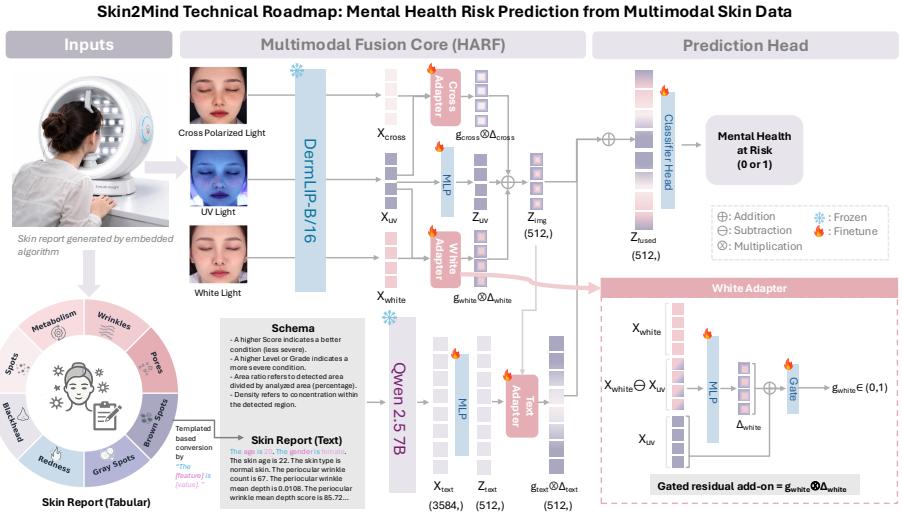


Fig. 1: **Overview of the Skin2Mind framework.** UV images serve as the primary anchor modality, while cross-polarized and white-light inputs contribute complementary information through gated residual pathways. Textual representations are integrated with the fused image anchor through gated additive fusion. Each adapter computes gated residual add-ons for auxiliary modalities; the architecture of the white adapter is illustrated, and the remaining adapters share the same design.

(1) To the best of our knowledge, Skin2Mind is the first computational framework that directly predicts individual-level, DSM-aligned mental health risk from quantitative facial skin phenotypes.

(2) We propose a novel HARF architecture. By establishing anchoring modalities and mitigating redundancy through gated residual pathways, HARF enables the stable integration of correlated clinical inputs.

(3) We introduce an interpretable, decision-aware LOFO analysis that provides clinically grounded cross-modal explanations.

2 Method

We propose Skin2Mind, a multimodal framework that integrates multi-spectral facial skin images and structured skin reports for mental health risk prediction (Fig. 1). The three visual inputs are denoted as X_{uv}, X_{cross}, and X_{white}, and the structured skin report as T. The pipeline consists of three stages. (1) **Unimodal Feature Extraction:** DermLIP-B/16 [22] encodes skin images, while Qwen2.5-7B [17] encodes the skin report into semantic representations. (2) **HARF:** Fusion follows a hierarchical anchor–residual design to reduce modality redundancy. UV images (X_{uv}) form the visual anchor and are refined by cross-polarized and

white-light inputs through gated residual adapters. The fused image representation is then used as the anchor for integration with textual features. (3) **Risk Prediction:** The resulting psychodermatological representation is fed into a lightweight classifier to predict binary mental health risk (at-risk vs. not-at-risk).

2.1 Unimodal Feature Extraction

Image Encoding. Facial skin images (X_{uv}, X_{cross}, X_{white}) are individually encoded using the vision encoder of DermLIP-B/16, a dermatology-specific vision–language foundation model [22]. The frozen vision encoder extracts L2-normalized features, which are linearly projected into 512-dimensional embedding vectors ($x_{\text{uv}}, x_{\text{cross}}, x_{\text{white}} \in R^{512}$).

Text Encoding. Skin analysis reports, originally in tabular form, are converted into plain-text sequences using a fixed template ("The [feature] is [value]") [10]. A scoring schema is prepended to the report to clarify the semantic meaning of quantitative attributes (e.g., "a higher score indicates a less severe condition"). The resulting text sequence is tokenized directly, without chat templates or instruction prompts. We use the frozen Qwen2.5-7B model (non-instruction-tuned) rather than the DermLIP text encoder, as long structured reports exceed CLIP-style short-caption scope. A fixed-dimensional embedding is obtained via mean pooling over the final-layer hidden states.

2.2 Hierarchical Anchor Residual Fusion

Our fusion design follows two anchor principles: (i) UV images serves as the visual anchor, and (ii) the fused image representation acts as the anchor for image-text integration. Auxiliary modalities are incorporated through gated additive refinements. The selection of UV as the visual anchor is supported by component-level ablations shown in Fig. 2(a), where UV consistently achieves the strongest predictive performance. The HARF design is motivated by the redundancy across modalities: multi-spectral images share similar structural content under different illumination, and skin reports are derived from the same imaging system, leading to partial visual correlations. Compared to naive concatenation, additive gating enables controlled information injection while preserving anchor stability.

UV as the Image Anchor. Given modality embeddings $x_{\text{uv}}, x_{\text{cross}}, x_{\text{white}} \in R^d$, we compute a UV anchor $z_{\text{uv}} = f_{\text{uv}}(x_{\text{uv}})$, where f_{uv} is a two-layer MLP with ReLU activation. For each auxiliary modality $m \in \{\text{cross}, \text{white}\}$, a residual feature $\Delta_m = h_m([x_m, x_{\text{uv}}, x_m - x_{\text{uv}}]) \in R^d$ is computed, where $[\cdot]$ denotes concatenation and h_m is a two-layer MLP with ReLU activation. A scalar gate $g_m = \sigma(\phi_m([x_{\text{uv}}, \Delta_m])) \in (0, 1)$ is applied, where ϕ_m is a two-layer MLP and $\sigma(\cdot)$ denotes the sigmoid activation function. The unified image representation is defined as:

$$z_{\text{img}} = z_{\text{uv}} + \sum_m g_m \, \Delta_m, \tag{1}$$

which preserves the dominant UV signal while selectively incorporating complementary illumination cues.

Image as the Multimodal Anchor. For cross-modal fusion, let $x_{\text{text}} \in R^{d_t}$ denote the text embedding. We project it into the visual space via $z_{\text{text}} = P_t(x_{\text{text}}) \in R^d$, where P_t is a two-layer MLP with LayerNorm, ReLU activation, and dropout. Although text and image embeddings carry overlapping content, they reside in heterogeneous representation spaces, so P_t maps text features directly without the explicit subtraction used for image-side anchoring. A scalar gate $g_{\text{text}} = \sigma(G_t([z_{\text{img}}, z_{\text{text}}])) \in (0, 1)$ is computed, where G_t is a two-layer MLP with sigmoid output. The final representation is obtained via:

$$z_{\text{fused}} = z_{\text{img}} + g_{\text{text}}\, z_{\text{text}}. \tag{2}$$

Optimization. The fused embedding z_{fused} is fed into a lightweight MLP classifier for risk prediction. The model is trained using Cross-Entropy (CE) loss with an additional sparsity regularization term on the fusion gates, where the expectation $E[g] = \frac{1}{B}\sum_{i=1}^{B} g_i$ denotes the mini-batch mean gate activation. We fix $\lambda = 10^{-3}$ in all experiments:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda\big(E[g_{\text{text}}] + E[g_{\text{cross}}] + E[g_{\text{white}}]\big). \tag{3}$$

2.3 Skin Report Semantic Attribution via ΔAUC

To evaluate semantic contributions, we group skin attributes into 34 clusters (e.g., pores, wrinkles, pigmentation) and perform cluster-level ablation via LOFO analysis. For each cluster c, related sentences are removed from the report and the model is retrained. The impact is quantified as $\Delta\text{AUC}_c = \text{AUC}^{\text{full}} - \text{AUC}_c^{\text{ablated}}$, where larger values indicate greater reliance on the removed semantic information. All training settings are kept identical.

3 Experiments

3.1 Dataset and Implementation Details

Dataset. We collected a dataset of 192 participants (51 not-at-risk, 141 at-risk) from the Sleep Medicine Center, West China Fourth Hospital, Sichuan University. The cohort included 148 males and 44 females, aged 15.5–76.7 years (mean 42.2). Multi-spectral facial images (cross-polarized, UV, white-light) and structured skin reports were acquired using the EveLab Insight-Eve M system (462 quantitative features spanning structural aging, pigmentation, and inflammation etc.). Mental health status was assessed using DSM-5-TR Level-1 Cross-Cutting Symptom questionnaires [1], covering multiple domains (e.g., depression, anxiety, mania, somatic symptoms, sleep disturbance, repetitive thoughts and behaviors, and substance use). Individuals exceeding the DSM-5-TR threshold for Level-2 assessment in any domain were labeled as at-risk, forming a binary classification task. Five-fold stratified cross-validation was used. Ethics approval and consent were obtained.

Implementation Details. The framework is implemented in PyTorch. For multimodal fusion, we employ an MLP classifier with two hidden layers (128 and 64 units), ReLU activations, and a dropout rate of 0.3. The model is optimized using AdamW with a constant learning rate of 1×10^{-3} and weight decay of 1×10^{-4}. We apply gradient clipping (max ℓ_2 norm of 1.0) and early stopping based on validation AUC (patience = 20) with a batch size of 32 for up to 500 epochs. Results are reported as mean ± standard deviation across cross-validation test folds. Code is available at https://github.com/skin2mind/miccai-2026.

3.2 Comparison with Baselines and State-of-the-Art Methods

We evaluate Skin2Mind against both unimodal and multimodal baselines (Table 1). Four out of six LLM encoders outperform the tabular XGBoost baseline on AUC, highlighting the benefit of semantic representation learning. General-domain LLMs (e.g., Mistral, Qwen) perform better than the domain-specialized BioMistral model, indicating that in frozen embedding setting, strong general-purpose backbones are sufficient for this task. In the image-only setting, the dermatology-pretrained DermLIP encoder outperforms ResNet (AUC 65.10% vs. 52.41%), supporting the importance of domain-adapted visual features. Significant differences appear mainly in AUC (which measures ranking across all thresholds), while F1 (at a fixed 0.5 threshold) shows fewer gaps, suggesting models differ more in risk ranking than in final binary predictions.

Table 1: **Comparison with Baselines and State-of-the-Arts.** Methods are grouped by modality and sorted by AUC within each group. **Bold** and underline denote the best and second-best results, respectively. Stars (*,** ,***) indicate statistical significance in AUC and F1 compared to our HARF-Qwen model at $p < 0.05, 0.01, 0.001$ levels respectively, determined by paired t-tests with Benjamini-Hochberg (BH) correction. Time is estimated from the wall-clock runtime of the embedding extraction loop (progress-bar runtime), measured on a single NVIDIA A100 80GB GPU, excluding data loading and preprocessing.

Modality Group	Method	Backbone	F1 % (mean ± std)	AUC % (mean ± std)	Time (ms)
Text-only (Tabular Skin)	Gradient Boosting	XGBoost	83.24 ± 1.76	67.17 ± 5.48**	/
	MLP	-	60.30 ± 21.60	66.90 ± 13.36	/
Text-only (Skin Report)	BioMistral-7B [13]	Mistral-7B	79.11 ± 2.55	53.31 ± 6.55***	53
	Deepseek-V2 [6]	Deepseek-7B	78.28 ± 3.75	53.93 ± 6.76***	380
	PubMedBERT [8]	BERT-base	80.31 ± 2.95	67.79 ± 5.23*	7
	LLaMa3.1-8B [20]	LLaMa3.1-8B	84.21 ± 3.08	69.59 ± 6.14*	529
	Qwen2.5-7B [17]	Qwen2.5-7B	81.08 ± 3.56	71.79 ± 6.12	518
	Mistral-7B [11]	Mistral-7B	<u>85.97 ± 4.64</u>	<u>78.41 ± 7.10</u>	588
Image-only (Facial)	ResNet-50 [9]	CNN	78.95 ± 2.33	52.41 ± 3.18**	416
	DermLIP-B/16 [22]	CLIP-ViT-B/16	79.53 ± 2.96	65.10 ± 4.33**	427
Text+Image (VLM)	Qwen3-8B-VL [2]	Qwen3-8B	82.93 ± 3.42	74.34 ± 3.86	3350
	Qwen2.5-VL [3]	Qwen2.5-7B	**87.45 ± 1.91**	75.31 ± 10.20	3740
Text+Image (Ours)	HARF-Mistral	Mistral-7B + DermLIP	86.01 ± 2.38	73.48 ± 4.31*	1015
	HARF-Qwen	Qwen2.5-7B + DermLIP	84.81 ± 1.52	**81.66 ± 4.32**	945

The advantage of multimodal integration lies in improved robustness, reduced performance variability, and computational efficiency. While strong text-only models achieve competitive AUC, they exhibit greater variability across folds. In contrast, HARF-Qwen achieves the highest mean AUC among all evaluated methods (81.66% ± 4.32%) and significantly outperforms most unimodal baselines. This suggests multimodal integration stabilizes discrimination by leveraging complementary signals from facial skin phenotypes and structured skin reports. Among end-to-end VLMs, HARF achieves the highest mean AUC (81.66%) with competitive variance (±4.32%) and F1 (84.81% ± 1.52%); Qwen3-VL is marginally more stable (±3.86%) but 7 points lower in mean AUC. Moreover, HARF's decoupled late-fusion architecture avoids the heavy inference cost of large-scale VLMs. HARF-Qwen achieves a per-sample latency of 945 ms, representing a 4× speedup over Qwen2.5-VL (3740 ms).

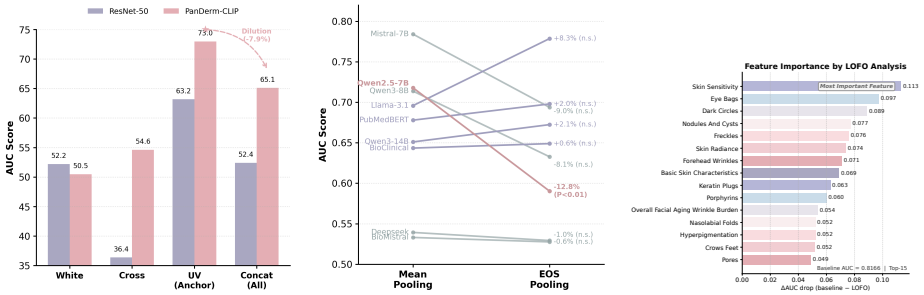
3.3 Ablation Studies

We conduct three ablation studies to analyze (1) fusion mechanism, (2) visual anchoring strategy and text representation, and (3) LLM model scaling.

Fusion Mechanism. Table 2 compares fusion strategies. The text-only model achieves an AUC of 71.79%, serving as a strong semantic baseline, while the tri-illumination image-only model reaches 65.10%. Naive fusion does not improve performance: Early Fusion-Concat achieves 69.17% AUC, while Late Fusion-Weighted (global softmax weights) and Late Fusion-Gated (sample-wise adaptive weights) further drop to 64.93% and 62.90%, indicating sensitivity to modality redundancy. Text-anchored fusion achieves 73.00% AUC (+1.21%), indicating modest gain, while our image-anchored residual fusion reaches **81.66%** AUC. This suggests text appears less suitable as the anchoring modality under redundancy, yielding smaller gains than UV-anchored fusion.

Table 2: **Ablation Study on Fusion Mechanism.** All models use fixed backbones (Qwen2.5-7B and DermLIP-B/16) to isolate the effect of fusion design. Naive fusion methods do not consistently improve over unimodal baselines, while anchor design plays a critical role in stabilizing multimodal learning.

Setting	Fusion Strategy	AUC (%)	Gap vs Text
Uni-Modal	Text Only	71.79	-
	Image Only	65.10	-6.69
Naive Fusion	Early Fusion-Concat	69.17	↓ 2.62
	Late Fusion-Weighted	64.93	↓ 6.86
	Late Fusion-Gated	62.90	↓ 8.89
Anchor Variant	Text-Anchored Fusion	73.00	↑ 1.21
Ours	**HARF**	**81.66**	↑ **9.87**



(a) Modality Selection (b) Pooling Strategy (c) Feature Importance

Fig. 2: **Model Ablations and Feature Importance.** (a) Performance across image illumination settings. (b) Comparison of text pooling strategies; (c) Top-15 influential features measured by ΔAUC drop (Baseline AUC = 0.8166). Statistical significance in (b) is assessed by paired t-tests with BH correction.

Visual Anchoring Strategy and Text Representation Figure 2(a) analyzes visual anchoring behaviour, showing that UV illumination provides the most stable and representative signal for the DermLIP encoder, while naive combination of all image modalities introduces feature dilution. Figure 2(b) compares text pooling strategies. Mean pooling yields slightly higher AUC values across a majority of backbones, although the differences are largely not statistically significant.

LLM Model Scaling (Text-Only Setting). We analyze model scaling under the *text-only* setting. Performance does not monotonically improve with model size: Qwen2.5-7B with mean pooling achieves the best result (71.79% AUC), outperforming larger instruction-tuned models such as Qwen3-8B (69.24%) and Qwen3-14B (71.24%). The Qwen3-8B embedding model performs worse (60.69% AUC), suggesting embedding-oriented representations may compress task-relevant clinical semantics.

3.4 Interpretability and Clinical Alignment

Figure 2(c) summarizes semantic-level LOFO attribution results. Larger ΔAUC drops indicate stronger contributions, as removing feature-related text leads to greater performance decline. As shown in Figure 2, the most influential features cluster into four dimensions closely associated with the skin-brain axis: epidermal barrier reactivity (Skin Sensitivity) [5], periorbital fatigue cues (Eye Bags, Dark Circles) [19], hypothalamic-pituitary-adrenal (HPA) axis-related pilosebaceous activity (Nodules and Cysts, Porphyrins) [18], and aging-related structural signals (Freckles, Forehead Wrinkles) [23]. These patterns indicate that the model relies on clinically interpretable dermatologic phenotypes that plausibly reflect broader neurophysiological and mental health mechanisms.

4 Conclusion

In this paper, we introduce Skin2Mind, to the best of our knowledge, the first psychodermatology-grounded computational framework that models the relationship between quantitative facial skin phenotypes and individual-level mental health risk. By integrating multi-spectral facial skin imaging with structured skin reports through a hierarchical anchor-based fusion strategy, the framework reduces modality redundancy, stabilizes cross-modal learning, and preserves clinically interpretable decision patterns. Experimental results demonstrate strong predictive performance with reduced variability and improved computational efficiency enabled by the decoupled design. Decision-aware LOFO analysis further identifies dermatologic dimensions consistent with skin-brain axis mechanisms, providing the first computational evidence that facial skin phenotypes may serve as scalable, non-invasive biomarkers for mental health risk screening. As an early exploration of this emerging direction, further evaluation of optimal model backbones, modality combinations, and validation in larger cohorts is warranted.

Acknowledgments. The authors thank the Sleep Medicine Center at West China Fourth Hospital, Sichuan University for clinical data collection. The study was approved by the Medical Ethics Committee of West China Fourth Hospital of Sichuan University (approval ID: HXSY-EC-2021022, HXSY-EC-2025133), and all participants provided written informed consent.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. American Psychiatric Association: Dsm-5-tr online assessment measures. https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/dsm-5-tr-online-assessment-measures (2022), accessed: 2026-02-26
2. Bai, S., Cai, Y., Chen, R., et al.: Qwen3-VL Technical Report (2025). https://doi.org/10.48550/ARXIV.2511.21631, https://arxiv.org/abs/2511.21631, version Number: 2
3. Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., Zhong, H., Zhu, Y., Yang, M., Li, Z., Wan, J., Wang, P., Ding, W., Fu, Z., Xu, Y., Ye, J., Zhang, X., Xie, T., Cheng, Z., Zhang, H., Yang, Z., Xu, H., Lin, J.: Qwen2.5-VL Technical Report (2025). https://doi.org/10.48550/ARXIV.2502.13923, https://arxiv.org/abs/2502.13923, version Number: 1
4. Bobok, N., Taskesen, T.: Stress-Induced Changes of the Skin: A Narrative Review. Cureus (Nov 2025). https://doi.org/10.7759/cureus.96285, https://www.cureus.com/articles/429743-stress-induced-changes-of-the-skin-a-narrative-review
5. Chen, Y., Lyga, J.: Brain-Skin Connection: Stress, Inflammation and Skin Aging. Inflammation & Allergy-Drug Targets **13**(3), 177–190 (Jun 2014). https://doi.org/10.2174/1871528113666140522104422, http://www.eurekaselect.com/openurl/content.php?genre=articleissn=1871-5281volume=13issue=3spage=177

6. DeepSeek-AI, Liu, A., Feng, B., Wang, B., et al.: DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model (2024). https://doi.org/10.48550/ARXIV.2405.04434, https://arxiv.org/abs/2405.04434, version Number: 5
7. El-Den, S., Chen, T.F., Gan, Y.L., Wong, E., O'Reilly, C.L.: The psychometric properties of depression screening tools in primary healthcare settings: A systematic review. Journal of Affective Disorders **225**, 503–522 (Jan 2018). https://doi.org/10.1016/j.jad.2017.08.060, https://linkinghub.elsevier.com/retrieve/pii/S0165032717313344
8. Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J., Poon, H.: Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. ACM Transactions on Computing for Healthcare **3**(1), 1–23 (Jan 2022). https://doi.org/10.1145/3458754, http://arxiv.org/abs/2007.15779, arXiv:2007.15779 [cs]
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition (2015). https://doi.org/10.48550/ARXIV.1512.03385, https://arxiv.org/abs/1512.03385, version Number: 1
10. Hegselmann, S., Buendia, A., Lang, H., Agrawal, M., Jiang, X., Sontag, D.: TabLLM: Few-shot Classification of Tabular Data with Large Language Models
11. Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., Casas, D.d.l., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L.R., Lachaux, M.A., Stock, P., Scao, T.L., Lavril, T., Wang, T., Lacroix, T., Sayed, W.E.: Mistral 7B (2023). https://doi.org/10.48550/ARXIV.2310.06825, https://arxiv.org/abs/2310.06825, version Number: 1
12. Koo, J., Lebwohl, A.: Psycho dermatology: the mind and skin connection. American Family Physician **64**(11), 1873–1878 (Dec 2001)
13. Labrak, Y., Bazoge, A., Morin, E., Gourraud, P.A., Rouvier, M., Dufour, R.: BioMistral: A Collection of Open-Source Pretrained Large Language Models for Medical Domains. In: Findings of the Association for Computational Linguistics ACL 2024. pp. 5848–5864. Association for Computational Linguistics, Bangkok, Thailand and virtual meeting (2024). https://doi.org/10.18653/v1/2024.findings-acl.348, https://aclanthology.org/2024.findings-acl.348
14. Mar, K., Rivers, J.K.: The Mind Body Connection in Dermatologic Conditions: A Literature Review. Journal of Cutaneous Medicine and Surgery **27**(6), 628–640 (Nov 2023). https://doi.org/10.1177/12034754231204295, http://journals.sagepub.com/doi/10.1177/12034754231204295
15. Merten, T.: The self-report fallacy: When diagnosis predominantly relies on subjective symptom report. Current Opinion in Psychology **65**, 102096 (Oct 2025). https://doi.org/10.1016/j.copsyc.2025.102096, https://linkinghub.elsevier.com/retrieve/pii/S2352250X25001095
16. Mulinari, S.: Aligning digital biomarker definitions in psychiatry with the National Institute of Mental Health Research Domain Criteria framework. NPP—Digital Psychiatry and Neuroscience **2**(1), 15 (Oct 2024). https://doi.org/10.1038/s44277-024-00017-6, https://www.nature.com/articles/s44277-024-00017-6
17. Qwen, Yang, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Li, C., Liu, D., Huang, F., Wei, H., Lin, H., Yang, J., Tu, J., Zhang, J., Yang, J., Yang, J., Zhou, J., Lin, J., Dang, K., Lu, K., Bao, K., Yang, K., Yu, L., Li, M., Xue, M., Zhang, P., Zhu, Q., Men, R., Lin, R., Li, T., Tang, T., Xia, T., Ren, X., Ren, X., Fan, Y., Su, Y., Zhang, Y., Wan, Y., Liu, Y., Cui, Z., Zhang, Z., Qiu, Z.: Qwen2.5 Technical Report (2024). https://doi.org/10.48550/ARXIV.2412.15115, https://arxiv.org/abs/2412.15115, version Number: 2

18. Saric-Bosanac, S., Clark, A.K., Sivamani, R.K., Shi, V.Y.: The role of hypothalamus-pituitary-adrenal (HPA)-like axis in inflammatory pilosebaceous disorders. Dermatology Online Journal **26**(2), 13030/qt8949296f (Feb 2020)
19. Sundelin, T., Lekander, M., Kecklund, G., Van Someren, E.J.W., Olsson, A., Axelsson, J.: Cues of fatigue: effects of sleep deprivation on facial appearance. Sleep **36**(9), 1355–1360 (Sep 2013). https://doi.org/10.5665/sleep.2964
20. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., Lample, G.: LLaMA: Open and Efficient Foundation Language Models (2023). https://doi.org/10.48550/ARXIV.2302.13971, https://arxiv.org/abs/2302.13971, version Number: 1
21. Vidal Yucha, S.E., Tamamoto, K.A., Kaplan, D.L.: The importance of the neuro-immuno-cutaneous system on human skin equivalent design. Cell Proliferation **52**(6), e12677 (Nov 2019). https://doi.org/10.1111/cpr.12677
22. Yan, S., Hu, M., Jiang, Y., Li, X., Fei, H., Tschandl, P., Kittler, H., Ge, Z.: Derm1M: A Million-scale Vision-Language Dataset Aligned with Clinical Ontology Knowledge for Dermatology (2025). https://doi.org/10.48550/ARXIV.2503.14911, https://arxiv.org/abs/2503.14911, version Number: 2
23. Yegorov, Y.E., Poznyak, A.V., Nikiforov, N.G., Sobenin, I.A., Orekhov, A.N.: The Link between Chronic Stress and Accelerated Aging. Biomedicines **8**(7), 198 (Jul 2020). https://doi.org/10.3390/biomedicines8070198