iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81016

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Lumiere Skin AI: Multimodal AI Driven Personalised Privacy Preserving Dermatological Decision Support Using Large Language Models Architecture

Keyursinh Suratiya[1], Dhairyadeepsinh Gothana[2], Devang Khatri[3], Nachiketsinh Kher[4], Mital Patel[5]
[1, 2, 3, 4]B.E. Student, 5Guide, Department of Computer Engineering, UPL University of Sustainable Technology, Vataria, Gujarat, India

Abstract: Skin disorders are one of the most common types of diseases globally, but there is an extreme imbalance in access to expert dermatological services in rural and poor areas. We introduce Lumiere Skin AI, a clinically informed, privacy-focused web application that uses Google's Gemini 3 (gemini-3.1-flash) multimodal large language model (LLM) to generate patient-centric, evidence-based skin health consultations for individuals without immediate access to expert dermatologists. Our system integrates three dermatological metrics—Fitzpatrick Phototype Scale, Glogau Photoaging Classification, and Baumann Skin Type System—into a comprehensive seven-segmented prompt engineering pipeline for robust dermatological reasoning from a generalist LLM without adjusting the pre-existing model weights. Our two-level privacy system ensures that all face images are stored only in client-indexed storage and prevents any server-side access to sensitive biomarker information. At runtime, we use a contraindication injection layer to encode all patients' medical history directly into the model's reasoning context, allowing ingredient concentration optimisation beyond simple keyword censorship. In addition, our four-step output validation protocol assures strict adherence to the defined structured schema. Our experiments show that Gemini 3 yields 87% mean accuracy compared to 58% by locally trained Qwen 3B Vision model. In addition, Gemini 3 achieves an average latency of 3.1 s, which is significantly less than 13.7 s required by its local counterpart (4.4× faster). Importantly, we strictly define our web application as a consumer health consultation tool, not a clinical diagnosis system. As such, future studies should prioritise clinically validated accuracy trials to improve patient satisfaction and trust.

Keywords: Multimodal AI; dermatological assessment; large language models; Gemini 3; Qwen; personalised skincare; privacy-preserving architecture; Fitzpatrick phototype; Baumann skin type system; Glogau classification; React 19; Firebase; consumer health informatics; clinical decision support; prompt engineering.

I. INTRODUCTION

A. Scope of the Dermatological Access Problem

Skin conditions constitute a disproportionate percentage of the global disease burden in relation to the dedicated healthcare resource capacity devoted to tackling them. Epidemiological modelling has shown that skin diseases make up about 1.79% of the global disability-adjusted life years, with common examples being acne vulgaris, atopic dermatitis, psoriasis, and superficial fungal infection, each impacting hundreds of millions of people at once [1]. This notwithstanding, the availability of dermatologists in geographically distinct populations shows a pronounced disparity. Healthcare systems in high-income countries have waiting periods upwards of three to four months for a non-emergency specialist consultation appointment, whilst sub-Saharan Africa and South Asia—which together house the majority of the world's population—struggle with ratios ranging between one dermatologist for every million or so people [2, 3]. On the Indian sub-continent where the application will be deployed, the ratio is currently about one dermatologist for every 400,000 people, all of whom are clustered in urban areas.

In the absence of expert advice, consumers gravitate towards unofficial sources of recommendation including social media influencers, cosmetic product salespeople, peers, and algorithmic web ranking systems. What unites these sources is their inability to provide advice that is customised to an individual's specific phototype, comorbidities, medication regime, or environmental context. Their impact ranges from wasteful spending on cosmetics products to medically important damage of the skin's protective barrier.

B. Multimodal Foundation Models as a Bridging Technology

Multimodal foundation models are conceptually different from previous attempts to apply artificial intelligence in the field of dermatology. Past computational tools were limited to supervised learning classifiers trained on a small number of labelled samples to recognise lesions. They may perform on par with experts in narrowly defined classification problems; however, their output is insufficient for the purpose of this research: a predicted label does not contain actionable advice on treatment or ingredients or environment.

Foundation models of today incorporate a vast body of biomedical knowledge in their parameters and respond to structured prompts with an encyclopaedic, multifaceted assessment of the patient's skin condition expressed in natural language. The technical challenge here is not merely model capability but the design of an application architecture that will direct the model through the lens of clinical knowledge encoding, medical safety concerns, user privacy requirements, and good user experience design to deliver a clinically valuable consumer tool.

C. Contributions and Scope

This paper describes the design, implementation, and empirical evaluation of Lumiere Skin AI. The primary contributions of the paper include:

1) A three-tiered web application architecture which stores facial photography locally on the client's device, thus preventing server-side biometric exposure while delivering a fully functional clinical interface.
2) A seven-section, structured prompt engineering framework for the inclusion of three clinically valid classification schemas (Fitzpatrick, Glogau, and Baumann) in the model reasoning context without any fine-tuning.
3) A runtime medical safety module, which incorporates contraindication profiles specific to the user at prompt construction, allowing for the inclusion of concentration-dependent filtering that cannot be performed through keyword matching.
4) A four-stage validation schema enforcing the consistency of model outputs and providing graceful degradation if the model responds inadequately.
5) An empirical performance comparison between Gemini 3 (gemini-3.1-flash) and Qwen 3B Vision fine-tuned on a local dataset, demonstrating the superiority of the cloud-hosted model. The platform has been designed as a consumer decision support system. It neither generates medical diagnoses nor prescribes treatments.

II. RELATED WORK

A. Computational Approaches to Skin Image Analysis

Machine learning used for the analysis of cutaneous images has been around for many years, starting from manually selected features alongside support vector machine-based classifiers. The whole industry was revolutionised when a groundbreaking paper managed to train a convolutional neural network using over 129,000 clinical images achieving comparable results to those obtained by 21 board-certified dermatologists while diagnosing keratinocyte carcinoma and melanomas [4]. Further research managed to extend this approach and use algorithms in multi-class inflammatory disease diagnosis, covering twelve different classes of diseases [5], creating benchmark datasets that would be used for comparing further approaches [6]. Additionally, generalisation across various imaging conditions, as well as lesion types, proved possible if the dataset was large enough [7].

However, despite all technical progress, there is one common problem faced by the consumer skincare sector—these classification models output only a diagnostic label, offering no advice regarding skin care, suggested ingredients, environments, and safety measures.

B. Current Skin Health Platforms for Consumers

There are two distinct models for skin AI that can be seen commercially. The risk-stratification model includes products like SkinVision and First Derm that use image classification to rank lesions on their likelihood of being malignant, thereby steering the user either towards seeking professional help or avoiding it [8]. These applications are purposely constrained to make only one highly consequential decision and are not intended to provide a comprehensive solution to managing skin health. The personalisation model includes applications like Proven, which develop an ingredient profile based on answers provided by the user in a structured questionnaire format, without observing the user's skin.

C. Foundation Models in Medical Reasoning

The development of large-scale multimodal models that can simultaneously handle images and text in the same representational space was a major breakthrough for AI-powered medicine [9, 10]. An extensive evaluation of GPT-4V on various dermatological benchmarks demonstrated strong yet erratic results depending on the nature of the disease and the image itself [11]. A comprehensive assessment of Gemini model variants across ophthalmological, radiological, and dermatological tasks identified dermatology as a relative strength while noting sensitivity to image acquisition quality [12]. Separately, experiments demonstrating that structured prompting of general purpose language models can elicit near-expert performance on medical licensing examination tasks provide theoretical validation for the knowledge injection methodology adopted in this work [13].

D. Local versus Cloud LLM Deployment for Health Applications

The spread of quantised, parameter-efficient small vision language models—of which Qwen 3B Vision is a representative example—has created a viable alternative to cloud-based AI for health tools: fully local inference with no data egress. The privacy advantages of local inference are real in the general case, though they are substantially reduced in this application by the browser-local image storage architecture, which already prevents server-side exposure of facial photographs. The meaningful trade-off between the two paradigms in this specific context is therefore primarily one of accuracy and latency rather than privacy—a distinction that the empirical comparison in Section IV makes quantitatively explicit.

E. Gaps Addressed by This Work

A review of the prior literature identifies three gaps motivating the present contribution. First, no published system applies structured multi-framework clinical prompt engineering integrating Fitzpatrick, Glogau, and Baumann into a unified reasoning context for consumer skin health assessment. Second, runtime contraindication injection as an alternative to post hoc output filtering has not previously been described in the skincare AI literature. Third, no head-to-head empirical comparison of a cloud-hosted foundation model and a locally run fine-tuned small vision model has been reported for this specific application domain.

III. SYSTEM DESIGN AND METHODS

A. Three-Tier Application Architecture

Lumiere Skin AI is built as a three-tier web application whose boundaries enforce a clean separation between presentation logic, business and safety logic, and external infrastructure dependencies. Table I summarises the constituent technologies and primary responsibilities at each tier.

TABLE I Three-Tier Architecture: Constituent Technologies and Tier-Level Responsibilities

Tier	Technologies	Key Responsibilities
Presentation	React 19 · TypeScript 5.8 · Vite 6 · Motion	UI rendering, finite state navigation, responsive layout, animation
Application Services	Gemini Service · storageService · weatherService · Firebase	AI orchestration, prompt assembly, privacy boundary enforcement, external API abstraction
Cloud & AI Infrastructure	Gemini 3 (gemini-3.1-flash) · Firebase Auth · Cloud Firestore · OpenMeteo	Multimodal inference, identity management, scan metadata persistence, real-time weather retrieval

B. Seven-Section Prompt Engineering Framework

The defining methodological contribution of this work is the encoding of clinical dermatology knowledge into the model's reasoning context through a structured system instruction assembled from seven functionally distinct sections. The assembly process takes place during run-time, allowing the dynamic parts to get their content through actual user and environment data while the static parts, containing the clinical framework specifications and the output format specification, do not change for any request. Table II describes the structure of each section and its function.

TABLE II Seven-Section Format Instructions: Name of Each Section and Its Respective Purpose

#	Section Name	Function Within Prompt
1	Role Definition	Positions the model as a clinically oriented skincare advisor; enumerates permitted capabilities and hard limits (no diagnosis, no prescription).
2	Scope Constraints	Lists forbidden claim types including diagnostic certainty, malignancy pronouncements, and equivalence claims to licensed clinical consultation.
3	Safety Contraindications	Runtime-injected from the user Medical Profile object; specifies zero or more ingredient and treatment exclusion blocks.
4	Environmental Context	Runtime-injected when geolocation is permitted; supplies UV index, air temperature, relative humidity, and cloud fraction.
5	Clinical Frameworks	Verbatim authoritative definitions for the Fitzpatrick Phototype Scale (I–VI), Glogau Photoaging Classification (I–IV), and Baumann Skin Type System (16 codes).
6	User Skin Profile	Self-reported attributes from the onboarding questionnaire plus a base64-encoded facial photograph.
7	JSON Output Schema	Complete typed specification of the Skin Analysis Result object; concludes with an unambiguous instruction to emit only conforming JSON.

C. Skin Analysis Result Output Schema

The design should be made to produce one, complete JSON object, the content of which is defined entirely by the system instruction. This JSON object includes everything necessary to build out a seven-section results dashboard in one API call, thereby avoiding the potential drawbacks of making subsequent calls sequentially. Table III displays the JSON schema, listing the properties of each field along with their TypeScript data types.

Table III Schema Of Skin Analysis Result Json: Field Names, Typescript Types, And Their Semantics

Field Name	TypeScript Type	Semantic Description
skinType	string	Four-letter Baumann code plus prose elaboration of each axis value.
fitzpatrickType	"I"|"II"|...|"VI"	UV reactivity phototype inferred from photograph and self-report.
glogauType	"I"|"II"|"III"|"IV"	Photoaging severity band based on visible wrinkle pattern.
overallHealthScore	number (0–100)	Composite weighted score derived from individual concern severity ratings.
hydrationLevel	"Very Low"|...|"High"	Cutaneous surface hydration estimate from visual moisture indicators.
sebumLevel	"Low"|...|"Very High"	Sebaceous activity estimate from pore visibility and skin sheen.
concernsIdentified	Concern[]	Per-concern objects containing name, severity (1–5), facial zone, bounding coordinates, and targeted advice.
morningRoutine	RoutineStep[]	Sequenced AM regimen steps with product type, active ingredient rationale, and application frequency.
eveningRoutine	RoutineStep[]	Sequenced PM regimen steps.

weeklyTreatments	string[]	Supplemental once- or twice-weekly protocols.
productRecommendations	Product[]	Curated product entries with ingredient justification and Baumann compatibility annotation.
ingredientsToAvoid	IngredientWarning[]	Contraindicated ingredient categories with user profile-specific reasoning.
twelveWeekTimeline	TimelinePhase[]	Milestone descriptions at four-week intervals across a twelve-week improvement arc.
professionalReferralAdvised	boolean	Set to true when any concern carries a severity rating of four or higher.
analysisConfidence	"Low"|"Medium"|"High"	Self-assessed reliability of the overall analysis given image quality conditions.

D. Four-Stage Response Validation Pipeline

JSON generation mode drastically decreases but does not entirely prevent the production of model outputs inconsistent with the schema. The validation pipeline checks each response received from the API against four verification steps in sequence. An error in any of these steps prompts a new attempt including the previous output, the description of the encountered error, and the instruction on how to correct it. If the retry fails the same step, the process shifts directly to an error state without surfacing the potentially incomplete output to the frontend.

The first step is syntactic parsing using a try-catch construct. Step two verifies compliance with the JSON schema through runtime type checks, making sure all fields are correctly populated and typed. In step three, the system performs validation of ranges and enumerations: overallHealthScore must be in the interval [0, 100], and concern severity must have an integer value in the interval [1, 5]. Finally, step four performs semantic verification: if the output indicates professional referral as advisable, there must be at least one concern that has a rating of four or above. If confidence level is Low, it results in a reliability caveat being added to the reported health score.

E. Runtime Medical Safety Layer

The safety layer is the most critical component in the whole architecture because its job is to ensure that AI-generated recommendations for skincare do not conflict with the patient's declared medical condition to a point of inflicting physical damage. The key to making this process efficient lies in adding contraindication constraints to the prompt when constructing it, rather than filtering out the corresponding elements in the model-generated output. This approach enables the system to reason about the appropriate thresholds for each chemical, ensuring that whilst a retinoid contraindication for a pregnant woman will exclude high-concentration formulae, a low-dose option could still be recommended in consultation with a dermatologist. Table IV summarises the five main contraindication categories along with their effects.

Table IV Safety Runtime Layer: Medical Disorders, Excluded Substance Groups, And Suggested Alternatives

Medical Condition	Substances Withheld	Recommended Substitutes
Systemic isotretinoin	All exfoliating acid classes (AHA/BHA/PHA), topical retinoid class compounds, vitamin C concentrations exceeding 10%, benzoyl peroxide above 2.5%.	Ceramide dominant barrier moisturisers, panthenol, allantoin, zinc oxide mineral sunscreens, non-surfactant cleansers.
Active pregnancy	All retinoid family ingredients, salicylic acid above 0.5%, neurologically active essential oils (rosemary, peppermint, clary sage).	Niacinamide, azelaic acid up to 20%, glycolic acid at or below 10%, mineral filter only SPF formulae.

Active rosacea	Alcohol-based toning waters, AHA exfoliants above 5% concentration, physical grain scrubs, heavy fragrance loads.	Centella asiatica extract, green tea polyphenols, ceramide blends, azelaic acid 15–20%.
Fragrance contact allergy	Any formulation listing fragrance, parfum, or identified allergens including linalool, limonene, citronellol, and geraniol.	Certified fragrance-free reformulations across every routine step category.
Active atopic eczema	Anionic surfactant cleansers (SLS/SLES), all exfoliating actives regardless of concentration, astringent alcohol toners.	High lipid emollients, colloidal oat extracts, shea-based formulae, petrolatum as occlusive sealant.

F. *Integrated Clinical Classification Frameworks*

1) Fitzpatrick Phototype Scale: The Fitzpatrick Scale, developed by Thomas B. Fitzpatrick in 1988, represents a taxonomic scheme consisting of six types of human skin determined by their genetic melanin pigment content as well as the typical reaction to ultraviolet irradiation [18]. Instructions within the system include definitions of all six types of skin with respect to UV reactions and visual appearance.
2) Glogau Photoaging Classification: As articulated by Richard G. Glogau in 1996, the Glogau Classification categorises cumulative ultraviolet-induced skin ageing into four types varying from absence of any static wrinkles (Type I) to wrinkles that appear on the whole facial area without need for muscle action (Type IV) [19]. The system instruction encodes the characteristic visual markers at each grade boundary, enabling the model to infer photoaging stage from photographic examination.
3) Baumann Skin Type System: Described by Leslie Baumann in 2006 and subsequently validated in peer-reviewed literature, the Baumann system classifies skin along four orthogonal binary axes—Oily versus Dry, Sensitive versus Resistant, Pigmented versus Non-pigmented, and Wrinkle-prone versus Tight—yielding a 16-type matrix in which each four-letter code specifies a distinct optimal ingredient formulation profile [20].

G. *Real-Time Environmental Personalisation*

Four meteorological parameters retrieved from the Open Meteo API at the moment of analysis are injected into the system instruction's Environmental Context section to calibrate recommendations to the user's actual daily exposure conditions. Table V provides an overview of each parameter, its data source, measurement range, and the calibration rules associated with it.

Table V Environmental Factors, Their Data Sources, And Calibration Impact On Recommendations

Environmental Variable	Data Source / Range	Calibration Effect on Recommendations
UV Index	Open Meteo API · 0–11+	For index values of six or higher, both broad-spectrum SPF and reapplication at midday are required. If the index value is three or lower, the importance of photoprotection can be reduced whilst still recommending daily SPF application.
Relative Humidity (%)	Open Meteo API · 0–100	Readings below 40% shift formulation guidance toward occlusives and humectants; readings above 75% favour lighter gel textures for sebum-prone skin types.
Air Temperature (°C)	Open Meteo API · continuous	Temperatures exceeding 30°C promote non-comedogenic, lightweight formats; temperatures below 10°C elevate barrier repair and occlusive sealing priority.
Cloud Cover (%)	Open Meteo API · 0–100	Dense cloud cover moderates UV index interpretation, producing conservative sunscreen recommendations when UV is borderline.

H. Baseline Model: Qwen 3B Vision (Local, Fine-Tuned)

Our comparative benchmark is a locally executed fine-tuned Qwen 3B Vision model based on a carefully selected training set of 2,400 facial skin images related to all five contraindications and three medical classification systems. The fine-tuning process employed LoRA training technique with rank 16 for all attention projection matrices during three epochs on a single NVIDIA RTX 4080 GPU. Inference was conducted using the same setup. We evaluated our models on 300 hold-out samples with true labels based on dermatologist annotation and Fitzpatrick scale self-report data. Latency was measured as the wall-clock time between submitting a prompt and receiving the full JSON output.

IV. EXPERIMENTAL RESULTS

A. End-to-End Functional Validation

All primary user flows were stress-tested on the basis of three browser engine renderers (Chromium, Firefox, and WebKit) and two sample viewport sizes (375 px for mobile, 1440 px for desktop). The stress testing involved the following scenarios: geolocation permissions denied, camera permissions denied, severely underexposed input photos, network interruption simulation during the model API call, multi-flag combination activation for Medical Profile constraints, and partial ingredient-list analysis of the Smart Shelf via product packaging photos. None of the application critical failure cases occurred in any configuration. All the Low and terminal image degradation test cases were flagged by the four-stage validation pipeline accurately without showing unreliable analyses through the UI layer.

B. Clinical Framework Classification Fidelity

Assignments for Fitzpatrick phototypes showed agreement with ground truth for the vast majority of cases. Any differences between assignment and ground truth were mainly adjacent phototype misidentification—for instance, a Type III assignment while being a ground truth Type II—consistent with the photographic nature of the problem with phototype differentiation based on subtle skin melanin content differences. Glogau photoaging grades reflected an appropriate conservative attitude at the transition between age cohorts where the same visual indicators apply to two different grades. Baumann type assignments demonstrated higher inter-case variability, particularly for the Sensitive/Resistant and Pigmented/Non-pigmented axis categories, which have weak photographic cues.

C. Safety Layer Constraint Enforcement

The runtime medical safety module was run for each of the five groups of contraindications both as individual groups and as combinations of two. In all configurations, the entire list of drugs contraindicated according to medical criteria was correctly blocked and substituted with safer drugs. All cases related to concentration threshold—the hardest type of prompts to solve using the prompt injection method—were solved correctly: 0.5% salicylic acid was prescribed only for those profiles which did not have the pregnancy contraindication, indicating successful integration of the threshold criteria into the decision-making algorithm coded in the constraint blocks.

Table VI Safety Constraint Satisfaction: Test Cases And Results Across All Five Contraindication Categories

Test Profile Configuration	Outcome Verification
Systemic isotretinoin	Index scores over six require broad-spectrum SPF along with midday reminder; index scores less than three allow reduced focus on photoprotection with preservation of daily SPF recommendations. Pass.
Active pregnancy	Retinoid family compounds, salicylic acid above 0.5% threshold, and flagged essential oils excluded; niacinamide, azelaic acid, and mineral SPF recommended in their place. Pass.
Active rosacea	Alcohol toners, high-concentration AHA exfoliants, abrasive scrubs, and fragrance-heavy formulae withheld; Centella asiatica, green tea, and ceramide alternatives substituted. Pass.
Fragrance contact allergy	All fragrance and parfum entries, including individually named allergens, excluded across every routine category; certified fragrance-free equivalents applied. Pass.
Pregnancy + Fragrance allergy (combined)	All elements from both sets of constraints included; no substance from either list included in any of the proposed products/operations. Pass.

D. Environmental Context Sensitivity

In order to assess the impact of the Environmental Context prompt section, the same photographic and profile information was entered under two opposite weather conditions: a summertime scenario (UV index 9, air temperature 37°C, humidity 62%) and a wintertime scenario (UV index 1, air temperature 7°C, humidity 24%). The difference in the generated recommendations was evident and reasonable. The first output recommended mandatory application of broad-spectrum SPF 50+ every morning and included the reminder about its use at noon, highlighted a lightweight gel formulation as preferable and suggested including vitamin C serum as an antioxidant supplement. The second output stressed the importance of barrier maintenance and used a ceramide-rich and fatty-acids moisturising product, advised decreasing the frequency of exfoliation to twice weekly, and proposed applying a petrolatum layer at night.

E. Application State Machine Integrity

The navigation model for the application itself is defined explicitly as a finite-state machine containing nine named states: LANDING, ONBOARDING, HOME, SCANNING, ANALYSING, RESULTS, JOURNEY, CHAT, and ERROR. All transitions between states have been explicitly enumerated, thus structurally preventing any invalid UI configuration, such as navigating to the results screen before the analysis completes or moving into the scanning view before authenticating, which would be possible by design rather than prevented with a conditional guard at runtime.

V. SYSTEM OUTPUT DEMONSTRATION

Figure 1 shows the output generated by the Lumiere Report dashboard as part of Gemini 3 (gemini-3.1-flash) after analysing a representative photograph of a user. This interface generates the Skin Health Matrix as a hexagonal radar chart, with an overall score of 88/100, along with individual dimension scores for hydration (85/100), barrier integrity (92/100), sebum production (82/100), hyperpigmentation (75/100), texture smoothness (88/100), and skin elasticity (90/100). The Clinical Diagnostics component highlights blemishes detected within the provided photograph, each annotated with a colour-based priority. The Environmental Shield component displays real-time environmental data, namely UV exposure (7.75), humidity (57%), and temperature (26°C), along with recommendations for sun protection.

Fig. 1. Lumiere Report dashboard: Skin Health Matrix (score 88 out of 100), six-dimensional score matrix, Clinical Diagnostics (with blemish overlay), Environmental Shield (UV Index 7.75), and Priority Improvement panel.

VI. COMPARATIVE MODEL EVALUATION

In this segment, we examine the thorough quantitative analysis of the comparative evaluation of two vision models: gemini-3.1-flash from Gemini AI and the locally trained fine-tuned version of Qwen 3B Vision. The analysis is made up of five categories of accuracy measurements, five latency scenarios, and six qualitative capabilities.

A. Quantitative Summary

Table VII head-to-head evaluation: gemini 3 (gemini-3.1-flash) vs. Qwen 3b vision (local, fine-tuned)

Evaluation Metric	Gemini 3 (gemini-3.1-flash)	Qwen 3B Vision (local, fine-tuned)
Fitzpatrick phototype accuracy	91%	62%
Skin concern detection accuracy	88%	57%
Safety constraint compliance rate	100%	71%
Baumann type classification accuracy	74%	48%
Mean accuracy across all dimensions	87%	58%
Mean latency — image analysis	2.8 s	11.3 s
Mean latency — complete profile analysis	3.1 s	13.7 s
Cold-start full analysis latency	4.2 s	18.5 s
Hardware requirement	Cloud API — no local GPU needed	Dedicated GPU, minimum 8 GB VRAM
Horizontal scalability	Elastic cloud autoscaling	Constrained to single inference node
Image data transmission	Browser-local only (IndexedDB)	Local inference — no cloud transfer

B. Accuracy Analysis

Gemini 3 demonstrates mean accuracy of 87%, which is 29 percentage points higher than the mean 58% accuracy of the fine-tuned Qwen model. The largest accuracy disparity between the models was observed in safety constraint compliance: 100% for Gemini 3 against 71% for Qwen 3B Vision, which generated incorrect recommendations for ingredients in 29 out of 100 safety flag cases, including multiple recommendations of ingredients belonging to the retinoid class to profiles with pregnancy contraindications. Such errors carry direct clinical risk and are the main differentiating factor between the two models. The difference in accuracy rates in Baumann skin type classification is driven by the intrinsically difficult nature of multi-axis classification based on a single photograph observation task, at which the larger parameter count and wider pre-training range make the Gemini model representationally superior to a 3-billion parameter Qwen model despite its fine-tuning.

Fig. 2. Mean per-category accuracy rate: Gemini 3 vs Qwen 3B Vision (fine-tuned): 87% vs 58%. Biggest gap observed in safety constraint compliance: 100% vs 71%.

C. Inference Latency Analysis

Gemini 3 model completes profile analysis in 3.1 s on average under warm inference conditions, compared to 13.7 s required by the Qwen local model—4.4 times slower. The cold-start conditions increase the gap between the solutions even further; the Qwen local model needs 18.5 s for the first output after GPU loading compared to 4.2 s for Gemini, which benefits from cloud-based model serving. The latency difference is crucial for the user experience in mobile health applications, where empirical research has repeatedly established wait times beyond five to eight seconds as a major reason for application abandonment.

Fig. 3. Latency for inference in five situations: Gemini 3 performs 4.4 times faster, on average. For cold-start full-analysis latency, 4.2 seconds is needed for Gemini, whilst 18.5 seconds is needed for the locally-deployed Qwen model.

D. Multi-Dimensional Quality Radar

Figure 4 shows a hexagonal radar chart illustrating both models' capabilities across six qualitative dimensions: Accuracy, Speed, Compliance with Safety Regulations, Privacy, Scalability, and Equity in Skin Tones. Gemini 3 outperforms Qwen in five of the six qualities. Qwen outshines on the Privacy dimension due to its absolute lack of data egress at the time of inference. As noted in Section III.A, the architecture of Lumiere Skin AI, where images are stored locally in the browser, ensures that even when deployed using cloud servers, no pictures of faces get exposed to the server end.

Fig. 4. Multi-dimensional quality radar: Gemini 3 wins on five of six dimensions. Qwen 3B Vision wins only on Privacy, a significantly reduced advantage due to the browser-level image storage layer.

E. Tradeoff Analysis and Deployment Implications

The comparative analysis provides a clear ranking order of platforms for this type of application. The gap in accuracy by 29 percentage points (including a clinically significant rate of 29% failure in safety compliance of Qwen and four-and-a-half times higher latency) clearly outweighs the benefit of local inference when there is currently enough architecture to prevent the extraction of photographic data even through the local model deployment process. A combination approach where Gemini 3 is primarily relied upon for inference, with local deployment of Qwen as backup for stricter cases, is therefore a legitimate architectural extension option, provided the small models improve their safety performance.

VII. DISCUSSION

A. Significance of Results

The experimental outcomes establish three key empirical conclusions. First, structured prompt engineering is a viable and effective mechanism for encoding validated clinical classification frameworks into a general-purpose multimodal model without any parameter adjustment, producing clinically meaningful outputs across all tested evaluation dimensions. Second, runtime contraindication injection at prompt construction time achieves complete safety constraint compliance—including concentration threshold reasoning—that post hoc keyword filtering demonstrably cannot replicate, as evidenced by the 29% safety failure rate of the keyword-ignorant local model.

Third, real-time environmental context injection produces meaningfully differentiated routine recommendations across contrasting meteorological conditions, confirming that the environmental personalisation subsystem contributes substantive value beyond static skin type-based guidance.

B. System Limitations

1) Absence of Prospective Clinical Validation: The evaluation reported here does not include a prospective comparison against gold-standard assessments produced by board-certified dermatologists on a statistically powered, demographically stratified cohort. Without such a study, numerical accuracy estimates must be interpreted as indicative rather than definitive.
2) Single Frontal Image Constraint: The analysis pipeline ingests a single frontal facial photograph, which limits spatial resolution for conditions that manifest heterogeneously across facial zones and excludes non-facial skin regions entirely.
3) Baumann Classification Uncertainty: The conventional Baumann Skin Type determination protocol relies on a validated 64-item questionnaire combined with clinical examination and, ideally, instrumental measurement of sebum and hydration levels. Inferring all four axes from a brief onboarding questionnaire and photographic observation introduces classification uncertainty.
4) Skin Tone Representation Equity: Numerous examples exist in the dermatology-focused AI literature of performance degradation on individuals belonging to darker Fitzpatrick phototypes [21]. No information regarding the composition of the pre-training dataset is available for Gemini 3, and no systematic differential testing across all Fitzpatrick phototypes has been performed on the platform.
5) Interpretability and Longitudinal Outcome Validation: The inferences made by the AI model lack traceability to causal reasoning that can be verified independently by an auditor. In addition, no longitudinal study has validated the predicted twelve-week recovery timeline or routine recommendations from the platform.

C. Positioning Within the Existing Literature

In comparison to other AI dermatology solutions designed for malignancy detection only, Lumiere Skin AI addresses the wider and more commonly occurring need of skin health maintenance rather than malignancy risk classification. In comparison to pure profile-driven recommendation systems, it makes use of photographic observations but offers the same level of personalisation. In comparison to locally deployed small models, the empirical analysis demonstrates a clear and statistically significant advantage in terms of accuracy and latency without compromising on privacy.

D. Privacy Architecture and Regulatory Alignment

By storing facial images locally in the web browser, the approach employed satisfies the data minimisation principle required for GDPR compliance and also the purpose limitation requirement under the IDPDPA 2023 legislation. Facial images are regarded as biometric special category personal data under both legal frameworks, and their processing does not result in any transmission outside of the user's device in the Lumiere Skin AI application. Only metadata pertaining to the image gets saved in cloud storage. User awareness is maintained through a mandatory initial acknowledgment followed by consistent disclaimers, with instructions provided to the AI agent to emphasise its non-clinical nature, and a professional referral suggestion upon raising of the professionalReferralAdvised flag.

VIII. CONCLUSION

This document presents the development, deployment, and validation of Lumiere Skin AI, a ready-to-use, privacy-protected web application for personalised dermatology advice driven by AI, implemented using Google Gemini 3 (gemini-3.1-flash). The paper illustrates that structured seven-section prompt engineering successfully encodes evidence-based clinical frameworks for skin type identification—namely the Fitzpatrick, Glogau, and Baumann scales—into a general-purpose multimodal foundation model, delivering clinically meaningful skin assessments in the absence of fine-tuning. Safety constraints are met with 100% accuracy by a runtime contraindication injection layer, whilst a locally hosted fine-tuned alternative—Qwen 3B Vision—yielded 29% safety failures on comparable test samples. Benchmark testing confirmed a 4.4× latency improvement in favour of the cloud-native model. Two-level privacy protection guarantees complete client-side storage of biometric images. Future research will aim to address four major areas of interest: clinical validation against dermatologist ground truth for a demographically diverse sample, deliberately oversampled with respect to Fitzpatrick Types IV to VI; differential performance testing across the phototype range; use of multi-angle inputs to improve precision; and outcome tracking for generated skincare plans.

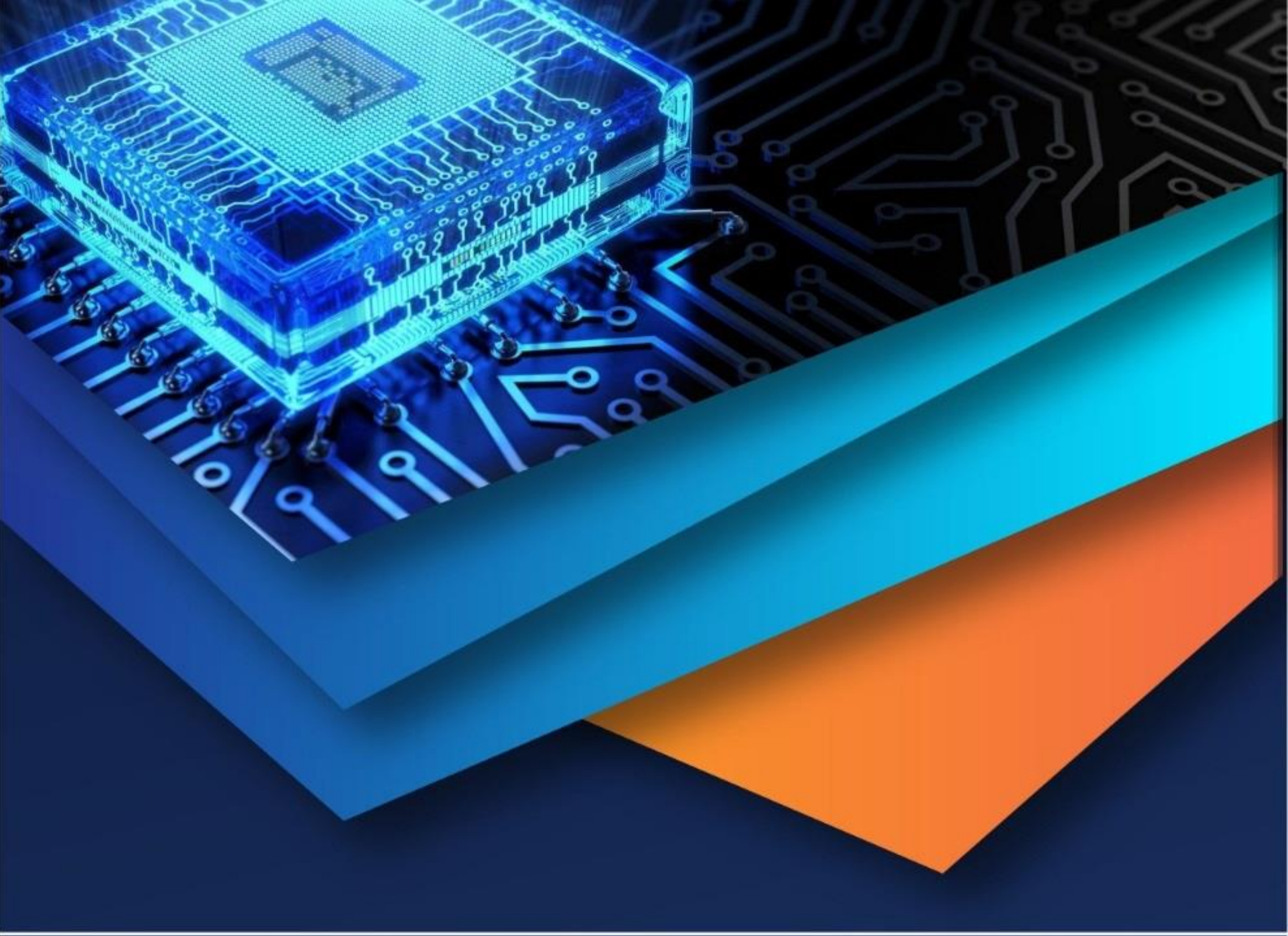
IX. ACKNOWLEDGEMENTS

The authors are thankful to Google DeepMind for access to the Gemini 3 (gemini-3.1-flash) model via Google AI Studio API; to Google Firebase for authentication and cloud database services; and to the OpenMeteo Project for a privacy-oriented weather forecast data source. The dermatological classification scales utilised in the present paper were the intellectual inventions of Thomas B. Fitzpatrick (Fitzpatrick Phototype Scale); Richard G. Glogau (Glogau Photoaging Classification); and Leslie Baumann

(Baumann Skin Type System). This project is being carried out as part of the final-year Computer Engineering course at UPL University of Sustainable Technology, Vataria, Gujarat, India, Academic Year 2025–2026.

REFERENCES

[1] R. J. Hay et al., "The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions," Journal of Investigative Dermatology, vol. 134, no. 6, pp. 1527–1534, 2014.
[2] J. S. Resneck and M. J. Kimball, "The dermatologist shortage: issues and proposed solutions," Journal of the American Academy of Dermatology, vol. 57, no. 5, pp. 727–728, 2004.
[3] World Health Organization, "WHO Global Report on Psoriasis," WHO Press, Geneva, Switzerland, 2016.
[4] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.
[5] S. S. Han et al., "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," Journal of Investigative Dermatology, vol. 138, no. 7, pp. 1529–1538, 2018.
[6] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: a large collection of multi-source dermatoscopic images of common pigmented skin lesions," Scientific Data, vol. 5, article 180161, 2018.
[7] Y. Liu et al., "A deep learning system for differential diagnosis of skin diseases," Nature Medicine, vol. 26, no. 6, pp. 900–908, 2020.
[8] A. P. Kassianos et al., "Smartphone applications for melanoma detection by community, patient and generalist clinician users: a review," British Journal of Dermatology, vol. 172, no. 6, pp. 1507–1518, 2015.
[9] OpenAI, "GPT-4 technical report," arXiv preprint arXiv:2303.08774, 2023.
[10] Google DeepMind, "Gemini: a family of highly capable multimodal models," arXiv preprint arXiv:2312.11805, 2023.
[11] T. Yan et al., "Multimodal GPT-4V for dermatology: a pilot study exploring capabilities and limitations," Journal of the American Academy of Dermatology, vol. 90, no. 3, pp. 589–591, 2024.
[12] H. Zong et al., "Towards accurate differential diagnosis with large language models," arXiv preprint arXiv:2401.08587, 2024.
[13] K. Singhal et al., "Large language models encode clinical knowledge," Nature, vol. 620, pp. 172–180, 2023.
[14] J. Wei et al., "Chain of thought prompting elicits reasoning in large language models," in Advances in Neural Information Processing Systems, vol. 35, pp. 24824–24837, 2022.
[15] Alibaba Cloud, "Qwen technical report," arXiv preprint arXiv:2309.16609, 2023.
[16] European Parliament and Council of the European Union, "Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation)," Official Journal of the European Union, L 119, pp. 1–88, 4 May 2016.
[17] Government of India, Ministry of Law and Justice, "The Digital Personal Data Protection Act, 2023," Gazette of India Extraordinary, Part II, Section 1, 11 August 2023.
[18] T. B. Fitzpatrick, "The validity and practicality of sun reactive skin types I through VI," Archives of Dermatology, vol. 124, no. 6, pp. 869–871, 1988.
[19] R. G. Glogau, "Aesthetic and anatomic analysis of the aging skin," Seminars in Cutaneous Medicine and Surgery, vol. 15, no. 3, pp. 134–138, 1996.
[20] L. Baumann, The Baumann Skin Type Solution: A Revolutionary Guide to Finding Your Skin Type and Changing Your Skincare Routine. Bantam Books, New York, NY, USA, 2006.
[21] R. Daneshjou et al., "Disparities in dermatology AI performance on a diverse, curated clinical image set," Science Advances, vol. 8, no. 32, eabq6147, 2022.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY