

A Novel Ensemble Learning Framework for Nonlinear Modeling of UV Exposure and Skin Damage

Resmi R¹, Sajan Jerome², Anu Babu³, Anilkumar E N⁴, Pramod P⁵, Baiju P S⁶

¹Professor LBS Institute of Technology for Women ,Thiruvananthapuram Dept.of Electronics and Communication Engineering <https://orcid.org/0000-0002-0849-5509>
(Corresponding author)

²Associate Professor LBS Institute of Technology for Women, Thiruvananthapuram Department: Dept.of Mechanical Engineering ORCID id : <https://orcid.org/0000-0003-4231-5265>

³Assistant Professor St Thomas Institute for Science and Technology, Trivandrum Dept.of Electronics and Communication Engineering ORCID id : <https://orcid.org/0000-0003-4304-5391>

⁴Professor LBS Institute of Technology for Women Dept.of Mechanical Engineering ORCID id : <https://orcid.org/0000-0001-5185-9489>

⁵Professor College name: LBS College of Engineering Kasaragod Dept.of Electronics and Communication Engineering In ORCID id :<https://orcid.org/0000-0002-1240-6548>

⁶Assistant Professor LBS College of Engineering, Kasaragod Dept.of Electronics and Communication Engineering ORCID id: <https://orcid.org/0009-0008-8332-6765>

Abstract

The Ultraviolet Index (UVI) is widely used to communicate the intensity of solar ultraviolet radiation, yet it remains a coarse public warning rather than a truly personalized risk indicator. The same ambient UVI can lead to very different biological outcomes once exposure duration, atmospheric conditions, and individual skin sensitivity are taken into account. This revised manuscript presents a structured framework for personalized UV risk prediction that combines multisource environmental sensing with nonlinear modeling and ensemble learning. At the core of the framework is a Skin Damage Index (SDI) that grows nonlinearly with UVI and exposure time, reflecting the fact that harm does not increase in a simple one-to-one manner at higher radiation levels. A fused prediction layer then integrates Random Forest, gradient boosting, and neural networks so that noisy tabular inputs, interaction-heavy patterns, and subtle nonlinear dependencies can be handled together rather than in isolation. The paper also introduces a modular architecture for data ingestion, preprocessing, inference, and user-facing recommendation delivery, making the framework suitable for implementation in mobile health platforms, wearable systems, and preventive dermatology applications. To improve interpretability, the manuscript includes system diagrams and analytical visualizations that show how risk evolves across different UV intensities and exposure durations. Overall, the work reframes UV safety from a single-number advisory problem into a personalized, context-aware prediction problem that is better aligned with real-world skin exposure.

Keywords: Ultraviolet Index, Skin Damage Index, Personalized UV Risk, Environmental Sensing, Ensemble Learning, Predictive Modeling, Dermatological Risk Assessment, Mobile Health.

1. Introduction

Ultraviolet exposure occupies a dual place in human health. In limited amounts it contributes to normal physiological processes, yet once intensity and duration begin to accumulate, the same radiation becomes a driver of erythema, photoaging, pigmentation changes, and long-term carcinogenic risk. Public awareness campaigns therefore rely heavily on the Ultraviolet Index, a simple scale that translates ambient solar UV intensity into an accessible warning. The UVI is helpful, but it is not the full story. It tells us what

the environment is doing; it does not directly tell us what a particular person is likely to experience on their skin.

That mismatch matters in practice. Two people outdoors under the same midday sun do not necessarily face the same biological burden. A fair-skinned user walking unshaded roads for forty minutes and a darker-skinned user stepping outside briefly are not exposed in the same way even if the ambient UVI is identical. Effective dose is shaped by duration, atmospheric transparency, reflective surroundings, altitude, behavior, and individual sensitivity. When those variables are ignored, a one-size-fits-all advisory can understate risk for some users while overstating it for others.

Recent progress in artificial intelligence, environmental sensing, and wearable health systems makes a more tailored approach feasible. Multisource data pipelines can combine satellite products, ground observations, and user-level inputs, while ensemble learning methods can represent complex interactions more flexibly than hand-crafted rules alone [1]-[7]. In parallel, data-driven systems in dermatology and medical decision support have demonstrated that nonlinear clinical patterns can be learned when the input space is thoughtfully structured [8]-[10]. This broader shift motivates the present work.

The contribution of this revised paper is not merely to restate the importance of UV awareness, but to reorganize the problem around personalized outcome prediction. Instead of asking only whether the sun is strong today, the framework asks a more useful question: given the ambient conditions, the likely duration of exposure, and the susceptibility of the user, how much skin stress is likely to accumulate and how should that risk be communicated?

Main contributions of the revised framework

1. A nonlinear Skin Damage Index is formulated so that high UVI values and longer exposure durations amplify risk in a biologically more realistic way than a linear approximation would allow.
2. Environmental, temporal, and user-specific variables are organized into a coherent feature stack rather than being treated as disconnected inputs.
3. An ensemble inference layer combines Random Forest, gradient boosting, and neural-network components so that robustness, discrimination power, and nonlinear interaction learning are preserved together.
4. The manuscript adds architecture diagrams, risk visualizations, and deployment logic so the method reads as an implementable system rather than a purely abstract model.

2. Research Gap and Design Objectives

Most existing consumer-facing UV advisories are intensity-centered rather than outcome-centered. They explain how strong the sun is, but they do not explicitly estimate how much biological stress a given person may accumulate. That gap is one reason public warnings can feel simultaneously useful and insufficient. The present framework is built around three limitations that recur in existing practice.

1. Linear thinking: many simple advisories imply that risk rises in near-proportion to UVI, even though real damage tends to accelerate once intensity and exposure time interact at the upper end of the scale.
2. Sparse personalization: ambient UV intensity is often communicated without enough regard for Fitzpatrick skin type, sensitivity, duration outdoors, or context-specific exposure modifiers.

3. Weak deployment logic: even when predictive ideas exist, they are not always translated into a modular pipeline that can support mobile alerts, wearables, or clinician-facing dashboards.

The design objectives of the proposed system follow directly from those shortcomings. A useful UV risk model should be personalized, nonlinear, explainable, and deployable. It should behave sensibly at both low and high UVI levels, tolerate noisy real-world data, and produce outputs that are easy to communicate in plain language. Just as importantly, it should remain extensible so that future clinical calibration, geographic adaptation, or integration with wearable biosensors can be added without redesigning the entire system [4], [7], [11].

3. Proposed UV Risk Prediction Framework

3.1 Data acquisition and fusion

The proposed framework draws on heterogeneous but complementary sources of information. At the environmental level, it can ingest satellite-derived UVI estimates, local ground-sensor readings, and meteorological descriptors such as temperature, ozone, aerosol load, and cloud-related attenuation. At the personal level, the model accepts exposure duration, skin-sensitivity indicators, and optional context flags such as outdoor activity pattern or protective coverage. The intention is to move from a single ambient signal to a richer representation of actual exposure opportunity.

Because those inputs are generated at different spatial and temporal resolutions, a data-fusion layer is required before prediction can begin. Timestamp alignment, unit normalization, quality filtering, and confidence scoring are therefore treated as first-class operations rather than afterthoughts. This matters because seemingly small inconsistencies, such as stale sensor readings or location mismatch, can distort downstream risk predictions more than model choice alone.

Table 1. Major feature groups used in the personalized UV risk framework

Feature group	Representative inputs	Preprocessing	Contribution to prediction
Environmental intensity	UVI, direct UV estimate, solar timing	Scaling and timestamp alignment	Provides the base irradiance burden
Atmospheric modifiers	Ozone, aerosol load, temperature, cloud effects	Normalization and quality filtering	Adjusts transmission and scattering of UV radiation
Exposure context	Expected outdoor duration, time of day, reflective surroundings	Interaction-term generation	Links ambient intensity to cumulative dose opportunity
User profile	Skin type, sensitivity history, optional coverage flags	Encoding and bounded scaling	Personalizes susceptibility to damage

Operational metadata	Location, sensor confidence, acquisition time	Traceability and missing-data handling	Improves reliability and interpretability of predictions
----------------------	---	--	--

3.2 Feature engineering and variable design

Raw inputs are not fed directly into the model. First, numerical variables are normalized so that models with different inductive biases can learn on compatible scales. Second, interaction terms are explicitly constructed for variables that are known to behave jointly, such as UVI $\times$ duration and UVI $\times$ skin sensitivity. Third, categorical attributes are encoded conservatively to preserve interpretability. These steps are simple in principle, but they help the learning system distinguish between a high ambient reading that lasts only a few minutes and a moderate reading that persists for a much longer period.

This feature-engineering stage also makes the framework resilient. In realistic deployment, some inputs will occasionally be absent or unreliable. By retaining confidence metadata and avoiding dependence on a single signal, the pipeline can degrade more gracefully instead of failing silently.

3.3 Nonlinear Skin Damage Index (SDI)

A central assumption of the manuscript is that biological damage should not be modeled as a purely linear function of ambient UV intensity. To reflect this, the framework defines an intermediate score called the Skin Damage Index. The SDI is designed to remain moderate at low UVI values, then rise more sharply as intensity and time accumulate.

$$SDI = \alpha \cdot (UVI)^\beta \cdot [1 + \ln(1 + t)] \cdot S \cdot E$$

In this formulation, α is a calibration constant, $\beta > 1$ captures nonlinear amplification at higher UVI levels, t represents exposure duration, S denotes user-specific skin sensitivity, and E is an environmental modifier summarizing context such as atmospheric transparency or reflective surroundings. The logarithmic term tempers the contribution of time at the lower end while still allowing prolonged exposure to accumulate risk. The value of β is especially important because it allows the system to represent the familiar but often under-modeled reality that a jump from UVI 9 to UVI 11 is not experienced in the same way as a jump from UVI 1 to UVI 3.

Once computed, SDI can be treated either as a continuous outcome for regression or as an intermediate score that is mapped to user-facing categories such as low, moderate, high, and very high risk. That flexibility makes the formulation useful for both analytical research and application-layer deployment.

3.4 Ensemble learning strategy

No single learning algorithm is ideal for all UV-risk scenarios. Tree-based models handle tabular heterogeneity well, boosting methods often sharpen discrimination around important thresholds, and neural networks can learn smoother nonlinear interactions that may be missed by rule-like partitions. The proposed solution is therefore an ensemble rather than a monolithic predictor.

$$\hat{y} = w_1 f_{RF}(X) + w_2 f_{GB}(X) + w_3 f_{NN}(X), \quad \text{where } w_1 + w_2 + w_3 = 1$$

Here, f_{RF} , f_{GB} , and f_{NN} denote the Random Forest, gradient-boosting, and neural-network components respectively, while the weights determine how strongly each model influences the final prediction. The ensemble can be configured through validation so that stable generalization is favored over any one model's local optimum. In operational settings, this often matters more than squeezing out a marginal gain on a single test split.

Table 2. Complementary roles of the ensemble components

Model component	Strength within the framework	Reason for inclusion
Random Forest	Stable with mixed-type and moderately noisy inputs	Provides a robust baseline and interpretable importance cues
Gradient Boosting	Captures high-order interactions and threshold effects	Improves discrimination when risk boundaries are tight
Neural Network	Learns smooth nonlinear couplings among features	Useful when exposure patterns are interaction-heavy or weakly separable
Weighted fusion	Combines complementary error profiles	Improves overall robustness relative to a single-model approach

3.5 Output classes and recommendation logic

The prediction layer is only useful if its output can be translated into clear guidance. For that reason, the system is designed to return both a continuous score and a categorical interpretation. A continuous score is valuable for researchers and system calibration; a categorical band is valuable for everyday users who need quick, understandable prompts.

The application layer can therefore issue different styles of feedback depending on the estimated risk and user context: a light reminder for low-to-moderate exposure, stronger warning messages when extended outdoor duration is expected, and time-sensitive alerts when conditions indicate a rapid transition toward harmful exposure. The same architecture can support phone notifications, wearable haptics, or clinician-facing review interfaces.

4. System Architecture and Data Flow

Figure 1 summarizes the operational architecture of the proposed system. The design is intentionally modular. Data collection and data fusion are separated from modeling so that future sources, such as wearable UV dosimeters or localized public-health feeds, can be integrated without rewriting the inference layer. Likewise, the nonlinear SDI block is kept explicit so that clinicians and engineers can inspect the damage formulation instead of treating the entire pipeline as a black box.

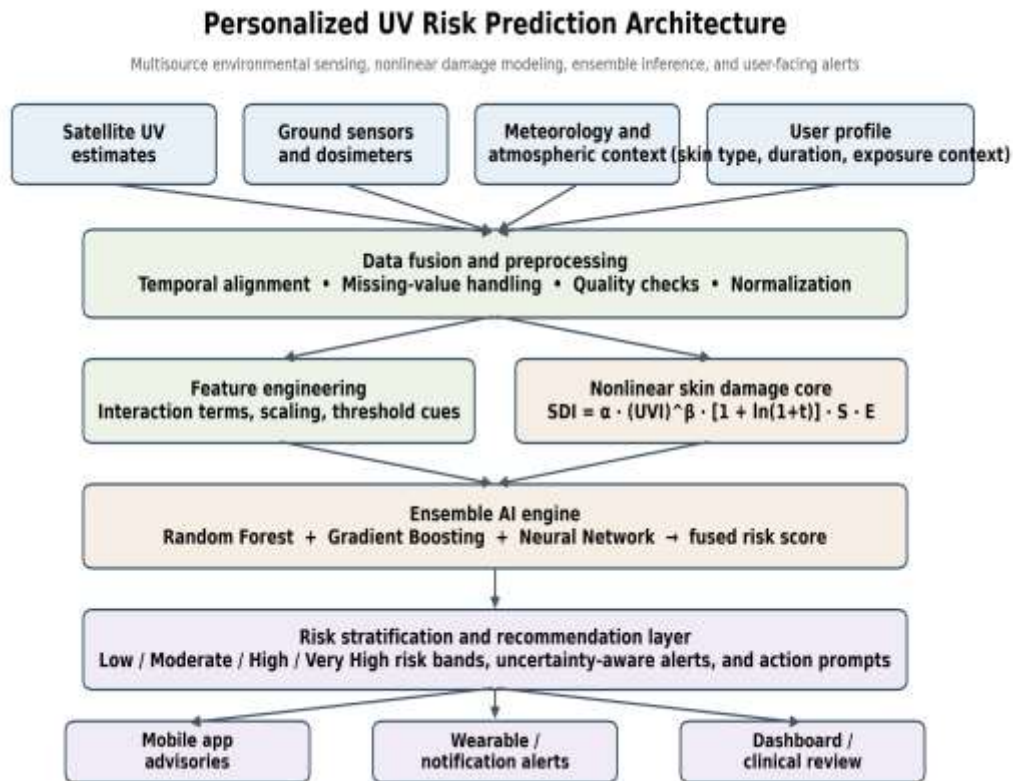


Figure 1. System architecture for multisource UV-risk prediction and user-facing alert delivery.

A key advantage of this architecture is that it supports more than one deployment pattern. A cloud-first version can perform data aggregation and inference centrally, whereas a lightweight mobile version can download environmental signals and run only the final scoring step locally. That flexibility matters if the framework is intended for public-health advisories, occupational monitoring, or personalized wellness applications across regions with different infrastructure constraints.

5. Illustrative Visual Analysis of UV Risk Behaviour

The visualizations in this section are analytical illustrations of the proposed model behaviour. They are included to explain how the framework responds to changes in UVI and exposure duration. In other words, they clarify the logic of the model rather than stand in for a completed population-scale clinical trial. This distinction is important because interpretability should be visible even before large-scale validation is complete.

5.1 Nonlinear interaction between UVI and duration

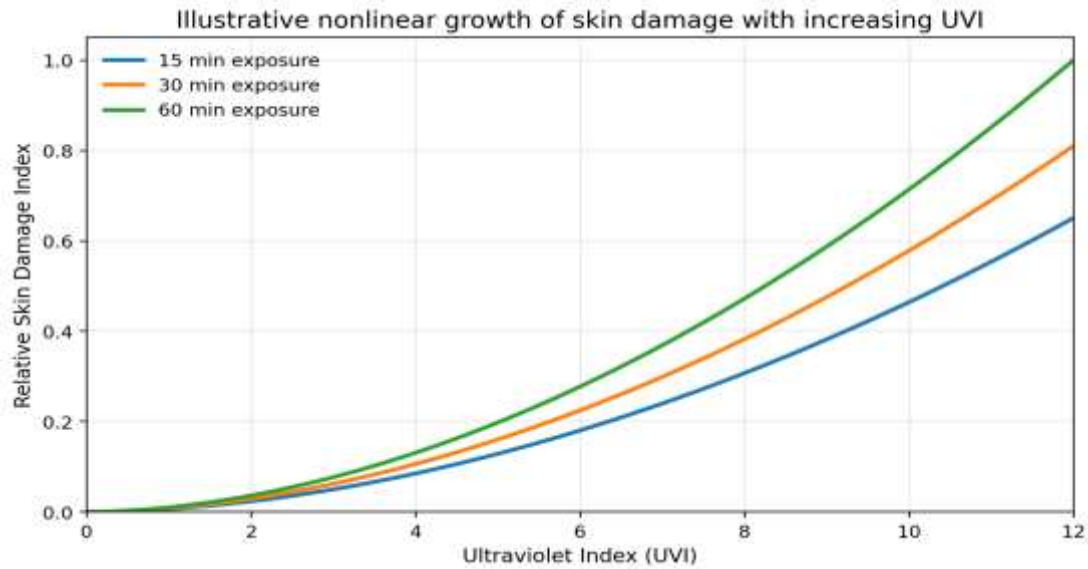


Figure 2. Relative Skin Damage Index under three exposure durations, illustrating the nonlinear response to increasing UVI.

The curves in Figure 2 show a pattern that simple linear warning systems often hide. At low UVI values the predicted difference between short and longer exposure windows remains noticeable but not dramatic. Once the index moves into the upper range, however, the gap between fifteen, thirty, and sixty minutes widens quickly. The user is not merely accumulating more of the same risk; the system is moving into a steeper biological regime.

This has practical implications. A forecast that seems only moderately worse than the previous day in terms of raw UVI may translate into a meaningfully larger increase in predicted skin stress when longer outdoor duration is also expected. That is precisely the type of interaction a nonlinear formulation is intended to retain.

5.2 Risk surface across UV intensity and exposure time

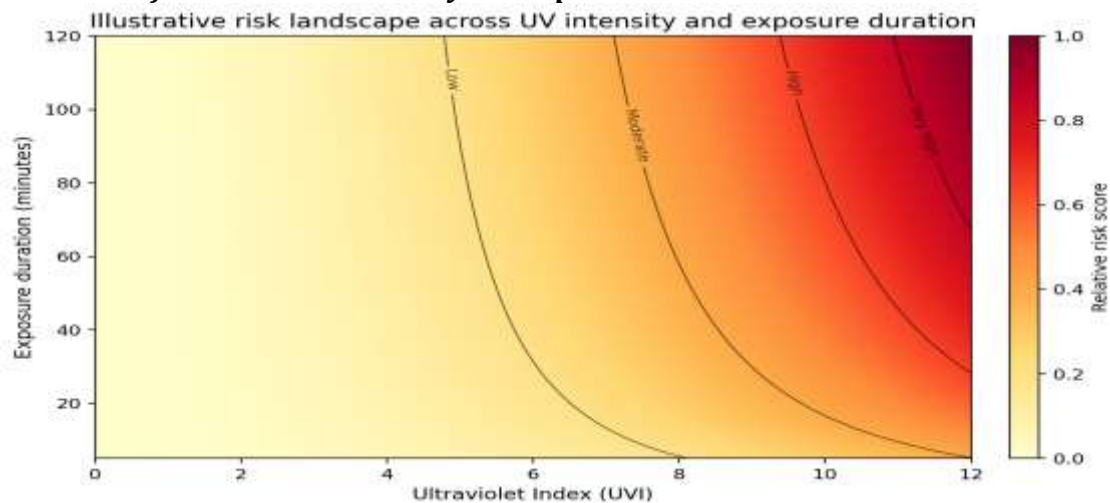


Figure 3. Risk surface showing how exposure duration and ambient UV intensity jointly shape the estimated risk score.

Figure 3 extends the same idea into a two-dimensional view. Short exposures under low UVI remain in the low-risk zone, but the surface bends upward as either duration or intensity increases. The curved transition bands are useful because they reveal where the system should move from passive advisory messaging to active warning. In a user application, those same bands can be translated into personalized messages or thresholds for notification timing.

A heatmap of this kind is also useful for calibration discussions with domain experts. Dermatologists, public-health practitioners, or occupational-safety researchers can inspect whether the transition from moderate to high risk occurs in a clinically sensible region and then refine thresholds accordingly.

Table 3. Example interpretation bands for application-layer communication

Risk band	Typical condition	Suggested response
Low	Brief exposure under low or mild UVI	Provide routine caution and remind the user to monitor duration.
Moderate	Extended time outdoors or brief exposure during stronger sunlight	Advise shade-seeking behaviour and protective measures before exposure accumulates.
High	Sustained outdoor activity during elevated UVI conditions	Issue an active alert and recommend reducing or rescheduling exposure where possible.
Very High	High UVI combined with prolonged exposure or high sensitivity	Trigger urgent warning, emphasize rapid protection, and communicate shorter safe-exposure windows.

6. Model Development and Evaluation Strategy

A full implementation of the framework would require paired inputs and outcome labels derived from carefully designed studies. Labels could come from dermatologist-reviewed erythema grading, controlled dosimeter readings, or structured self-report paired with environmental measurements. The essential point is that the model should learn from biologically meaningful endpoints, not from ambient UVI alone.

For continuous SDI estimation, error metrics such as MAE and RMSE are appropriate. For categorical risk prediction, accuracy, precision, recall, F1-score, and area-under-curve measures should all be reported. Just as important, probability calibration should be checked so that a predicted high-risk score is not merely separable but numerically trustworthy. In safety-oriented applications, calibration and subgroup reliability are often as important as headline accuracy.

Validation should also be stratified across skin-sensitivity groups and usage contexts. A model that performs well on average but underpredicts for a subset of users may be unacceptable in practice. Similarly, overprediction can reduce trust if the system becomes too alarmist. For that reason, subgroup audits, error analysis, and threshold tuning belong to the core evaluation protocol rather than to a post hoc appendix.

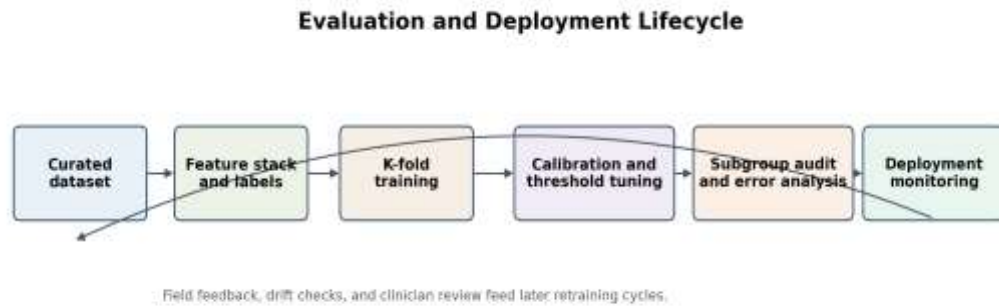


Figure 4. Recommended lifecycle for training, calibration, subgroup auditing, and post-deployment monitoring.

Once a validation cohort is available, standard diagnostic visualizations such as confusion matrices, calibration curves, and ROC plots can be generated to complement the architectural and analytical figures shown earlier. Interpretability tools, including feature-importance analysis or post hoc explanation methods, can then help verify whether the model is relying on sensible cues rather than on accidental correlations [6], [7], [10].

7. Practical Deployment Considerations, Ethics, and Limitations

Several real-world issues should be acknowledged before a system of this kind is presented as decision-ready. First, UV exposure is influenced by behaviour in ways that are difficult to capture perfectly. Clothing choice, shade availability, body orientation, sweating, movement between indoor and outdoor spaces, and user adherence to notifications can all affect the eventual biological outcome. Medications associated with photosensitivity may also shift risk in ways that a generic model does not fully represent.

Second, data quality and fairness require deliberate attention. Environmental measurements may be noisy, delayed, or spatially coarse; wearable streams can be incomplete; and user-level sensitivity variables may be simplified representations of a much richer clinical reality. Any serious deployment should therefore express uncertainty, not just a point prediction. In other words, the system should be honest about what it knows and what it is estimating indirectly.

Third, privacy and governance cannot be treated as optional. A personalized UV-risk service may rely on location traces, wearable histories, and behavioural patterns. Those data are meaningful, but they are also sensitive. Storage policies, consent design, retention limits, and transparency around inference logic become central if the framework is to evolve into a consumer or clinical product.

Finally, the present manuscript remains a framework paper rather than a completed clinical validation study. Its strength lies in the clarity of the proposed modeling and deployment logic. Its limitation is that the full system still depends on prospective validation, subgroup calibration, and long-horizon field testing before it can be interpreted as a medical or regulatory-grade decision tool.

8. Conclusion and Future Scope

This revised manuscript argues that UV safety should be modeled as a personalized and nonlinear prediction problem rather than as a single ambient warning signal. By combining a structured Skin Damage Index, multisource environmental data, and an

ensemble learning layer, the proposed framework offers a more realistic basis for estimating how UV exposure may translate into skin stress for different users.

The added diagrams and visual analyses make the system easier to interpret, while the deployment-oriented architecture clarifies how the framework could support mobile health applications, wearables, and preventive dermatology workflows. Future work should focus on prospective data collection, clinically grounded calibration, fairness auditing across skin categories, and adaptive recommendation strategies that evolve with user behaviour over time. With those steps in place, AI-assisted UV-risk prediction could become a practical tool for everyday prevention rather than a purely academic abstraction.

References

- [1] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019.
- [2] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.
- [3] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018.
- [4] H. Banaee, M. U. Ahmed, and A. Loutfi, "Data mining for wearable sensors in healthcare," *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1749–1759, 2013.
- [5] R. Esteva, A. Robicquet, B. Ramsundar, et al., "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, pp. 24–29, 2019.
- [6] G. Yang, Q. Ye, and J. Xia, "Artificial intelligence in biomedical signal and image processing: advances and challenges," *IEEE Signal Processing Magazine*, vol. 39, no. 2, pp. 82–97, 2022.
- [7] S. Azizi, A. Khosravi, and S. Nahavandi, "AI-driven healthcare: emerging technologies and challenges," *IEEE Reviews in Biomedical Engineering*, vol. 15, pp. 123–142, 2022.
- [8] D. Ting, L. Pasquale, L. Peng, et al., "Artificial intelligence and deep learning in ophthalmology," *Nature Medicine*, vol. 25, pp. 965–973, 2019.
- [9] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.
- [10] S. Saria, A. Butte, and A. Sheikh, "Better medicine through machine learning: what's real, and what's artificial?" *PLoS Medicine*, vol. 15, no. 12, 2018.
- [11] L. A. Celi et al., "The MIMIC-III critical care database," *Scientific Data*, vol. 3, 2016.
- [12] M. M. Warriar, L. Abraham, and R. Resmi, "An Optimal Deep Ensemble Model for the Classification of Pneumonia from Chest X-Rays," in *Proceedings of the 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2023, pp. 1–7.