МРНТИ 77.03.03 DOI: 10.62931/2959-6335_2026_1_19

A MOBILE APPLICATION FOR PERSONALIZED SELECTION OF SKINCARE PRODUCTS BASED ON CONVOLUTIONAL NEURAL NETWORKS AND VISION-LANGUAGE MODELS

Zhakupbekova D. B., Martyntsov N. V.[a]
Astana IT University, Astana, Kazakhstan

Corresponding author: Martyntsov N. V. Nikolay.Martyntsov@astanait.edu.kz

Abstract. This paper presents the architecture and software implementation of a mobile application for skin analysis and personalized skincare product selection driven by artificial intelligence techniques. The solution primarily targets individuals whose skin undergoes increased physiological stress – professional athletes and active fitness enthusiasts – while remaining applicable to a broader audience. The analysis module utilizes a two-stage pipeline: a ResNet-family convolutional neural network (CNN) trained via transfer learning on open-source selfie image datasets with dermatological labeling, and the Gemini Flash 2.5 vision-language model (VLM), which validates CNN outputs for complex classes. The recommendation algorithm is detailed, featuring stringent allergen filters, product scoring based on skin type, an active ingredient incompatibility graph, and selection criteria that account for price tiers and brand diversity. Quantitative performance metrics demonstrate an accuracy of 88.65% for skin type classification and 91.35% for skin concerns; the integration of the VLM validator improves accuracy on challenging class pairs from 79.0% to 92.4%. These results confirm the practical viability of the hybrid CNN + VLM approach for mobile facial analysis in cosmetic applications.

Keywords: skin analysis, convolutional neural networks (CNN), vision-language models (VLM), recommendation systems, skincare products, mobile applications, cosmetology.

Introduction. The skin of individuals who exercise regularly is exposed to factors that distinguish its condition substantially from that of an average consumer. Intense training is accompanied by prolonged sweating, pH shifts, mechanical friction from sports equipment, exposure to chlorinated or seawater, high-intensity ultraviolet radiation, and temperature fluctuations. These factors lead to the development of specific conditions: mechanical acne, chronic dehydration, regular microtrauma to the lipid barrier, and post-inflammatory hyperpigmentation.

Due to a chronically compromised protective barrier, an athlete's skin becomes hypersensitive. Under such conditions, the use of mass-market cosmetics without regard for individual skin characteristics, or the application of conflicting active ingredients (such as acids and retinoids), is not only ineffective but can also lead to severe contact dermatitis. This necessitates an exceptionally precise, dermatologically grounded approach to skincare selection – one that prioritizes allergen screening and ingredient compatibility, which generic retail recommendations fail to deliver.

Solving the problem of objective product selection is of high importance: for the general user, it reduces economic losses from ineffective purchases (according to Skin Trust Club surveys, 63% of buyers purchase products that do not match the needs of their skin [6]), while for the athletic audience it minimizes the risk of aggravating skin pathologies under heavy physical strain. The global beauty-tech market, according to Grand View Research, reached USD 66.16 billion in 2024 and is projected to grow to USD 172.99 billion by 2030 [5]. However, existing solutions (SkinVision [3], Haut.AI [7], Perfect Corp. AI Skin Diagnostic [8], TroveSkin [9]) are predominantly oriented toward neoplasm screening or are tied to a single brand's catalog and ignore the most important aspect – individual component intolerance and incompatibility between products.

The aim of this research is to develop a cross-platform mobile application that performs automatic facial skin condition analysis from a photograph and generates a safe, personalized skincare plan. The scientific novelty lies in the application of a two-stage hybrid “CNN + VLM” [4] pipeline to improve recognition accuracy of complex dermatological patterns on a mobile client, as well as in the formalization of an active ingredient incompatibility graph and a product selection algorithm that considers the correlation between the user's profile and a product, ensuring safe skincare even for skin subjected to extreme athletic stress.

Materials and methods. The system is built on a three-tier client-server architecture (Figure 1). The mobile client is implemented in Flutter with the Dart language, providing a single codebase for iOS and Android and performance comparable to native. The server side is built on FastAPI with Python 3.12 and asynchronous request handling; the average JSON request processing time is approximately 11 ms, with a throughput of 15,000–20,000 RPS, which is 3–5 times higher than comparable metrics for Django and Flask. Data is stored in PostgreSQL 17 with the pg_trgm extension for searching INCI compositions and JSONB fields containing ML analysis results.

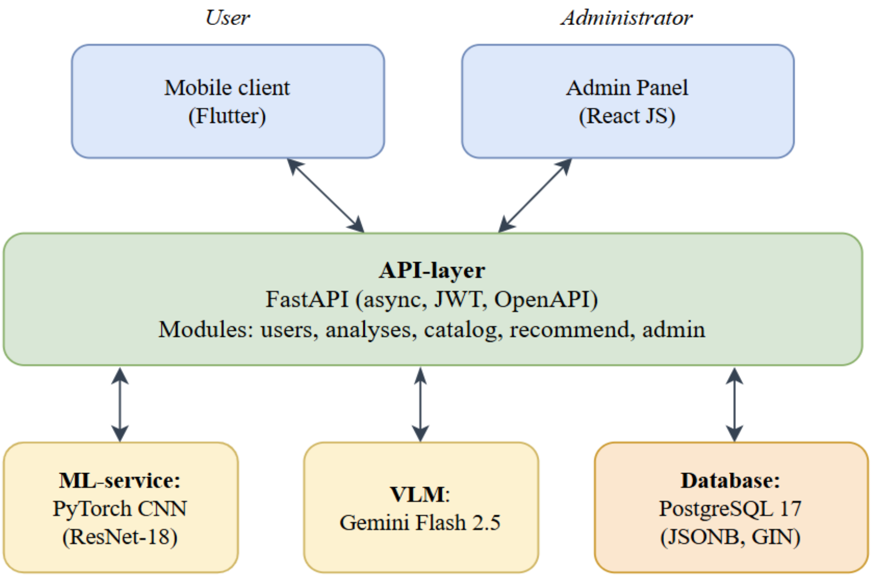

Figure 1. System architecture

The analysis module is based on two convolutional neural networks: the first classifies skin type (dry, normal, oily), and the second classifies detected concerns (acne, under-eye bags, redness, wrinkles, pigmentation, spots, scars, and others). To train the network, we assembled a dataset of approximately 9,500 images. It included both professional medical images from the open HAM10000 dataset [2] (photos exhibiting signs of pigmentation and scarring were used) and specially collected and labeled selfie photographs. The dataset was divided into three parts: 70% of the images were used for training, 20% for validation, and 10% for final independent testing. To enable the network to recognize faces better under varying conditions, all photos were standardized (224×224 px resolution, color normalization according to the ImageNet standard). In addition, augmentation was applied – artificially expanding the dataset through random transformations of images: horizontal flipping, rotations (up to ±15°), cropping, and adjustments of brightness and contrast (by ±20%).

The ResNet-18 neural network architecture, pretrained on ImageNet, was chosen as the base model due to its optimal balance between accuracy and runtime performance on mobile devices: the model has 11.7 million parameters, requires 1.8 GFLOPs of computational power, processes a single photo on a CPU in 220 ms, and achieves an accuracy of 91.3% (top-1). Alternative options were also tested (Table 1).

Table 1 – Comparison of models for the skin-condition classification task

Model architecture	Number of parameters, M	Inference time (CPU), ms	Accuracy (top-1), %	Rationale
ResNet-18	11.7	220	91.3	Selected as the baseline. Optimal balance between runtime performance and recognition accuracy.
ResNet-50	25.6	480	92.8	Rejected. A twofold increase in processing time and model size for a marginal accuracy gain (+1.5 percentage points).
EfficientNet-B0	5.3	160	90.4	Rejected. High variance (instability) in recognizing underrepresented (rare) skin concerns.
VGG16	138.0	1200	–	Rejected. Excessive parameter counts and critically low speed for a mobile client.
MobileNetV3-L	5.4	95	89.7	Reserve option. Considered as a candidate for future versions of the application with fully on-device operation.

Note: the dash in the accuracy column for VGG16 indicates that the model was rejected at the computational-complexity evaluation stage, before full accuracy testing on the dataset.

To adapt the model to the task, its final layer was modified: a dropout regularization mechanism (Dropout 0.5) and a new linear layer for skin-concern classification were added. Training proceeded in two stages (transfer learning) – for the first 20 epochs only the new final layer was trained, after which two deep blocks of the network were "unfrozen" and trained for an additional 10 epochs [1]. To prevent the network from ignoring minority classes, a specialized loss function was applied (weighted cross-entropy with inverse-frequency weights), which penalizes the algorithm for errors on rare classes. Training was carried out using the Google Colab cloud service based on Jupyter Notebook. The process was controlled by the AdamW optimizer (learning rate $1 \cdot 10^{-4}$, weight decay $1 \cdot 10^{-4}$) with smooth learning rate scheduling (CosineAnnealingLR). Images were fed in batches of 32. To avoid overfitting, the algorithm automatically stopped (early stopping) if its metrics did not improve over 5 cycles (patience = 5).

To minimize errors in recognizing visually similar skin concerns (for example, acne/redness or pigmentation/scars), a hybrid two-stage image-processing pipeline has been implemented in the application (Figure 2).

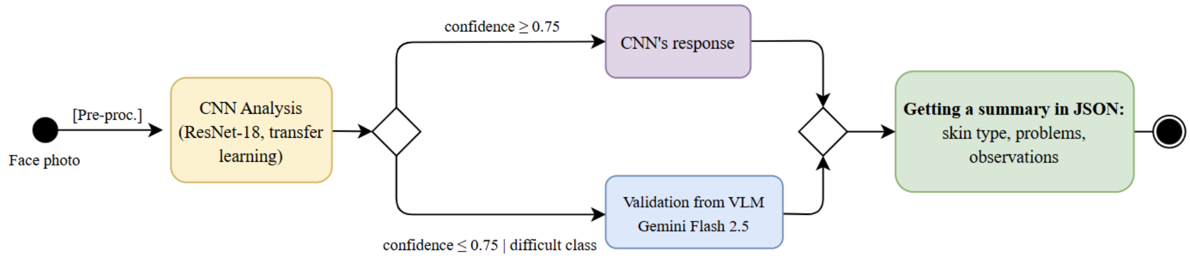

Figure 2. Two-stage facial image analysis pipeline

The algorithm produces a preliminary diagnosis and estimates its confidence in it. A confidence threshold of 75% (0.75) was determined empirically based on the best balance between precision and recall (maximum F1 score). If the base model is confident in its decision at ≥ 0.75, the pipeline terminates and immediately returns the result. If the confidence is below this threshold, or if the predicted concern belongs to the list of "complex" diagnoses, the system automatically launches the second stage – in-depth validation using the VLM Gemini Flash 2.5. In this case, the VLM receives the source photograph, the candidate diagnoses from the first network, and a structured prompt describing the classes. The model operates in zero-shot mode (without the need for additional fine-tuning): it analyzes the image, corrects the diagnosis, and produces the final response in JSON. As output, the system generates a structured JSON summary containing the skin type, identified concerns, and secondary multimodal metrics evaluated exclusively by the VLM due to its ability to analyze fine-grained skin texture. These metrics include numerical estimates for skin barrier function, specifically a hydration index (hydration_level ranging from 0 to 1) and an enlarged pores index.

An important aspect of deploying such an architecture is the balance between cost and legal safety. A single VLM call (analysis of a 1024×1024 px photo and generation of a 500-token response) costs USD 0.0016 [10]. At the same time, the use of free API versions (Free Tier) is categorically inadmissible when working with real biometric data (facial photographs). For this reason, the production version of the pipeline uses the enterprise Vertex AI gateway. Its terms (Section 17, Service Specific Terms) legally guarantee full encryption of data in transit and at rest, as well as a strict prohibition on using user photographs to train Google's neural networks [11]. This makes the diagnostic process fully confidential and ensures the application's compliance with the international GDPR standard [13] and the Law of the Republic of Kazakhstan “On Personal Data and Its Protection” [12].

Results. Testing the developed algorithm on an independent dataset confirmed its effectiveness. In the basic skin-type identification task, the overall accuracy was 88.65%. The algorithm recognizes oily skin with the highest confidence (F1 score = 0.91), while for dry and normal skin this metric was 0.89 and 0.86, respectively.

On the more challenging task of classifying specific concerns, the network achieved an overall accuracy of 91.35% (Figure 3).

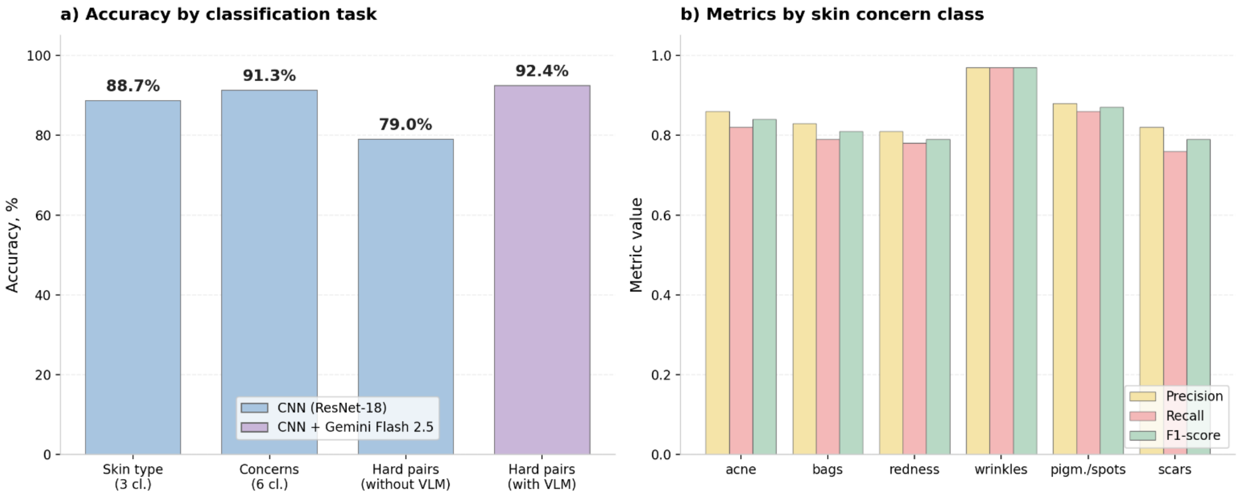

Figure 3. Quantitative evaluation of skin-analysis models

A detailed breakdown showed that the system most accurately detects wrinkles (F1 = 0.97). This is explained by the clear visual contours of this concern the substantial volume of evaluation data (≥ 300 examples in the test sample). The algorithm encountered the greatest difficulties in recognizing redness and scars – the F1 score for both classes dropped to 0.79.

The reasons for these difficulties are revealed by the confusion matrix of the classifier (Figure 4).

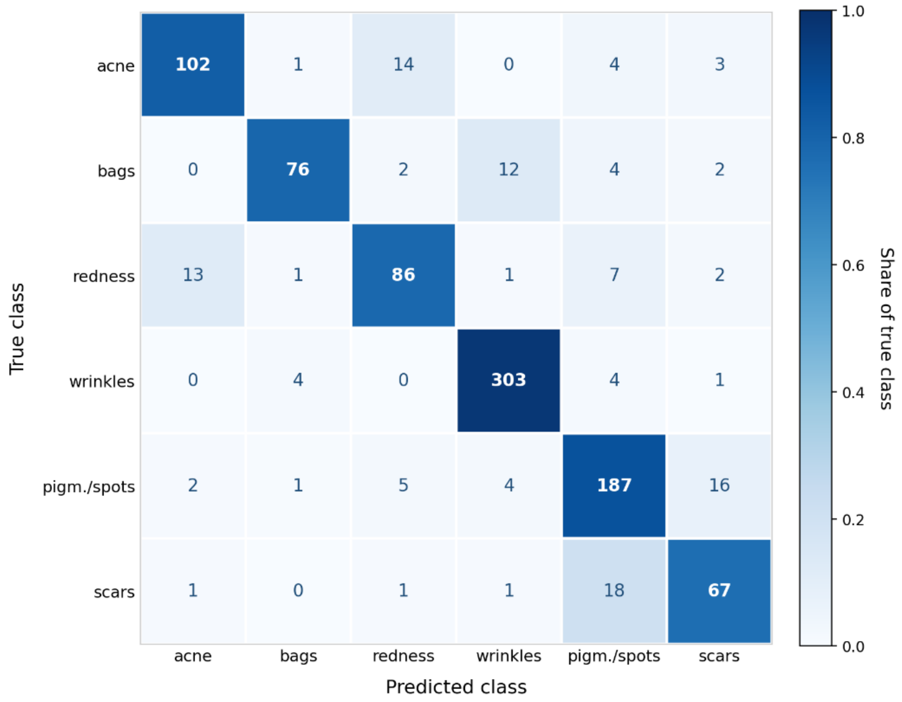

Figure 4. Confusion matrix of the skin-concern classifier

Its analysis revealed two systematic issues related to the visual similarity of different skin conditions:

In 11% of cases, the base network confuses acne with ordinary skin redness, since both of these conditions are accompanied by inflammatory color change (erythema).

In 20% of cases, the algorithm mistakes scars for pigmentation spots due to the visual similarity between skin depressions (atrophic scars) and post-inflammatory hyperpigmentation.

The second analysis stage (VLM) is used specifically to eliminate these “blind spots”. To evaluate its effectiveness, a separate dataset of 240 challenging photographs corresponding to the problematic pairs “acne/redness” and “pigmentation/scars” was assembled. Use of the hybrid approach made it possible to improve recognition accuracy on these complex images from 79.0% to 92.4%.

This solution is also economically viable. Because the system invokes the VLM not for every image, but only when confidence in the diagnosis is low (when the threshold filter is triggered), only 15–20% of all user photographs are sent to the resource-intensive second stage. This makes it possible to keep the average server-side cost of processing a single request low, rendering the architecture suitable for large-scale commercial use.

The skincare routine generation is implemented as a multi-stage pipeline (Figure 5). The input consists of a combined user profile (age, sex, allergies, free-form comment) and the JSON result of the analysis. Each product in the catalog passes through six stages.

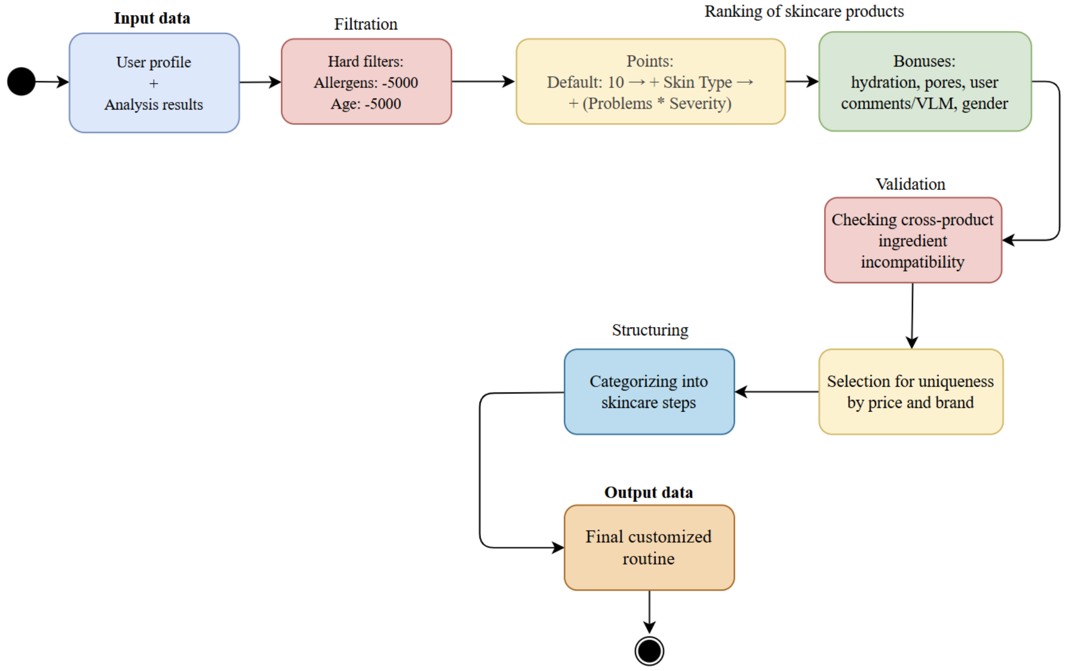

Figure 5. Pipeline for generating personalized recommendations

In the first stage, strict safety filters are applied. If a product is unsuitable for the user by age or if a declared allergen is found in its composition, the product receives a critical penalty (−5000 points), which guarantees its exclusion from the recommendations. To prevent the system from missing components due to typos or OCR errors on labels, composition comparison is performed using a fuzzy-matching algorithm with a similarity threshold of 0.85.

The second and third stages handle the base scoring. Each product is given a starting score of 10. An exact match with the user's skin type awards a bonus of +40 points, universality awards +5, and a mismatch deducts 15 points. Points are then added for addressing specific facial concerns: +20 points if the concern is declared in the product description, and +15 points for the presence of proven active ingredients (for example, salicylic acid and niacinamide for acne; peptides and retinol for wrinkles; vitamin C for pigmentation). To ensure the algorithm prioritizes the most pressing concerns, these points are multiplied by a dynamic coefficient according to the formula:

$$\frac{0.5 + severity}{75}, \qquad (1)$$

where severity of concern is expressed as a percentage.

The fourth stage is fine-tuning that takes into account additional observations from the VLM model and biological sex. If the model detects dehydration (hydration level < 0.5), the algorithm awards additional points according to the formula

$$+15 \times (1 - hydration\_level)$$

for the presence of hyaluronic acid, ceramides, and squalane. In the case of enlarged pores (index > 0.5), the presence of zinc and kaolin is rewarded. The gender filter takes physiology into account: for women experiencing acne and redness, the weight of soothing actives (centella, niacinamide) is increased by a coefficient of 1.3, which compensates for hormonal fluctuations in sebum production during the luteal phase of the cycle. For men, the same coefficient of 1.3 is applied to healing components (aloe, panthenol) to address microtrauma from shaving; and for oily male skin, priority for lightweight textures is increased by a factor of 1.2.

The fifth stage is a check for chemical conflicts via the active-ingredient incompatibility graph (Figure 6). This function is critically important for sensitive and damaged skin. The graph formalizes strict dermatological rules: for example, retinol cannot be combined with AHA/BHA acids and vitamin C due to differences in optimal pH and the risk of severe irritation; acids degrade peptides, while peptides accelerate the unwanted oxidation of vitamin C. The exception is niacinamide, which the algorithm recognizes as universal and compatible with all actives. In the implementation, the graph is a scalable static dictionary, CONFLICTING_ACTIVES, which makes it easy to update the rules when new popular actives appear on the market.

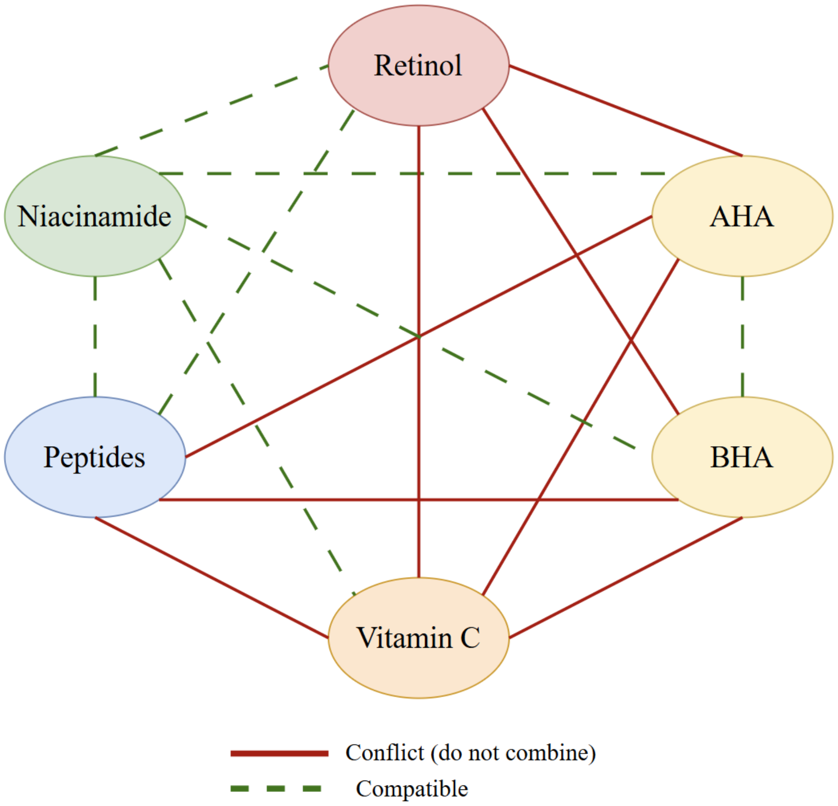

Figure 6. Incompatibility graph of active cosmetic ingredients

The sixth stage produces the final selection using a three-pass algorithm that ensures diversity of choice. On the first pass, the system selects one best product in each of three price tiers (budget, mid-range, premium), giving preference to different brands. On the second pass, it fills the list up to five items based on maximum score, and the third pass serves as a fallback in case of a narrow selection of products suitable for the user. The resulting products are distributed across 3–8 basic skincare steps (from cleansing to SPF).

The reliability of the entire pipeline, including verification of the incompatibility graph and the recommendation business logic, is ensured by automated testing. The code is covered by unit and integration tests based on the pytest framework. To isolate the test environment, interaction with the PostgreSQL database is implemented through the testcontainers-python library, while calls to the external neural network are emulated using a mock endpoint of the Gemini API.

Conclusion. In the course of this research, a cross-platform application for intelligent cosmetic product selection was developed. The effectiveness of the hybrid “CNN + VLM” pipeline has been demonstrated: the base ResNet-18 model classifies skin type with an accuracy of 88.65% and concerns with an accuracy of 91.35%, while the Gemini Flash 2.5 validator reduces the number of errors on difficult class pairs, raising accuracy from 79.0% to 92.4%.

A multifactor recommendation algorithm has been developed that excludes allergens, accounts for gender-related characteristics, and checks for chemical conflicts between products via the active-ingredient incompatibility graph. The software implementation is based on the Flutter, FastAPI, and PostgreSQL stack.

The practical value of this work lies in the creation of a safe skincare-selection tool, which is critically important for the target audience – athletes whose skin lipid barrier is regularly compromised. Prospects for further development of the system include extending the algorithms to body skincare selection, integration with services that track physiological indicators of the body, and the introduction of comprehensive medical support: assigning a personal dermatologist to each user and connecting allied specialists (gastroenterologists and endocrinologists).

МОБИЛЬНОЕ ПРИЛОЖЕНИЕ ДЛЯ ПЕРСОНАЛИЗИРОВАННОГО ПОДБОРА СРЕДСТВ ПО УХОДУ ЗА КОЖЕЙ НА ОСНОВЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ И VISION-LANGUAGE МОДЕЛЕЙ

Жакупбекова Д. Б., Мартынцов Н. В.[a]
Astana IT University, г. Астана, Казахстан

Автор для корреспонденции: Мартынцов Н. В. Nikolay.Martyntsov@astanait.edu.kz

Аннотация. В данной статье представлены архитектура и программная реализация мобильного приложения для анализа состояния кожи и персонализированного подбора средств ухода с использованием методов искусственного интеллекта. Решение ориентировано преимущественно на лиц, кожа которых подвергается повышенным физиологическим нагрузкам, – профессиональных спортсменов и людей, активно занимающихся фитнесом, – при этом приложение может использоваться и более широкой аудиторией. Модуль анализа реализован на основе двухэтапного подхода: сверточной нейронной сети (CNN) семейства ResNet, обученной методом transfer learning на открытых наборах селфи-изображений с дерматологической разметкой, и vision-language модели Gemini Flash 2.5, выполняющей валидацию результатов CNN для сложных классов. Подробно описан алгоритм рекомендаций, включающий строгие фильтры аллергенов, оценку продуктов в зависимости от типа кожи, граф несовместимости активных ингредиентов, а также критерии отбора с учетом ценовых категорий и разнообразия брендов. Количественные показатели эффективности демонстрируют точность 88,65 % при классификации типа кожи и 91,35 % при определении кожных проблем; интеграция VLM-валидатора повышает точность на сложных классах с 79,0 % до 92,4 %. Полученные результаты подтверждают практическую применимость гибридного подхода CNN + VLM для мобильного анализа лица в задачах косметологии.

Ключевые слова: анализ кожи, сверточные нейронные сети (CNN), vision-language модели (VLM), рекомендательные системы, средства ухода за кожей, мобильные приложения, косметология.

СВЕРТКАЛЫ НЕЙРОНДЫ ЖЕЛІЛЕР ЖӘНЕ VISION-LANGUAGE МОДЕЛЬДЕР НЕГІЗІНДЕ ТЕРІ КҮТІМІ ҚҰРАЛДАРЫН ЖЕКЕЛЕНДІРІП ТАҢДАУҒА АРНАЛҒАН МОБИЛЬДІ ҚОСЫМША

Жақыпбекова Д. Б., Мартынцов Н. В.[a]
Astana IT University, Астана қ., Қазақстан

Хат алмасу үшін автор: Мартынцов Н. В Nikolay.Martyntsov@astanait.edu.kz

Аңдатпа. Бұл мақалада жасанды интеллект әдістеріне негізделген тері жағдайын талдау және күтім құралдарын жекелендіріп таңдауға арналған мобильді қосымшаның архитектурасы мен бағдарламалық жүзеге асырылуы ұсынылады. Шешім, ең алдымен, терісі жоғары физиологиялық жүктемеге ұшырайтын тұлғаларға – кәсіби спортшылар мен белсенді фитнес әуесқойларына – бағытталған, сонымен қатар кең аудиторияға да қолдануға жарамды. Талдау модулі екі кезеңді тәсіл негізінде құрылған: дерматологиялық таңбаланған ашық селфи-суреттер деректер жиынтығында transfer learning әдісімен оқытылған ResNet тобына жататын сверткалы нейронды желі (CNN) және күрделі кластар үшін CNN нәтижелерін валидациялайтын Gemini Flash 2.5 vision-language моделі. Ұсыным алгоритмі егжей-тегжейлі сипатталған, оған аллергендерді қатаң сүзгілеу, тері түріне байланысты өнімдерді бағалау, белсенді ингредиенттердің үйлесімсіздік графы, сондай-ақ баға санаттары мен брендтердің әртүрлілігін ескеретін іріктеу критерийлері кіреді. Сандық тиімділік көрсеткіштері тері түрін жіктеуде 88,65 % және тері мәселелерін анықтауда 91,35 % дәлдікті көрсетеді; VLM-валидаторын интеграциялау күрделі кластардағы дәлдікті 79,0 %-дан 92,4 %-ға дейін арттырды. Алынған нәтижелер косметология саласындағы мобильді бет талдауы үшін CNN + VLM гибридті тәсілінің практикалық тиімділігін растайды.

Түйін сөздер: тері талдауы, сверткалы нейронды желілер (CNN), vision-language модельдері (VLM), ұсыным жүйелері, тері күтімі құралдары, мобильді қосымшалар, косметология.

References

1 He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – P. 770–778.

2 Tschandl P., Rosendahl C., Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions // Scientific Data. – 2018. – Vol. 5. – Art. 180161.

3 Bayaneni S. SkinVision: Leveraging Deep Learning for Skin Condition Analysis // Medium. – 2025. – URL: https://medium.com/@samanvitha9/skinvision-leveraging-deep-learning-for-skin-condition-analysis-6860d9c5add2.

4 Zeng W., Wang K. et al. MM-Skin: A Vision-Language Foundation Model and Dataset for Dermatology // arXiv preprint. – 2025. – arXiv:2505.06152. – URL: https://arxiv.org/abs/2505.06152.

5 Beauty Tech Market Size, Share & Trends Analysis Report By Product, By Distribution Channel, By Region, And Segment Forecasts, 2023–2030 // Grand View Research. – 2023. – URL: https://www.grandviewresearch.com/industry-analysis/beauty-tech-market-report.

6 Wray J. Nearly two thirds of women are wrong about their skin type // Cosmetics Business. – 2022. – URL: https://cosmeticsbusiness.com/nearly-two-thirds-of-women-are-wrong-about-their-skin-type-200155.

7 Haut.AI: official website. – URL: https://haut.ai/.

8 Perfect Corp. AI Skin Analysis: Skincare Solutions: official website. – URL: https://www.perfectcorp.com/business/showcase/skincare/home.

9 TroveSkin Demo: [application demo version]. – URL: https://isnur.github.io/troveskin-demo/.

10 Gemini 2.5 Flash Model Card and Pricing // Google DeepMind. – URL: https://ai.google.dev/gemini-api/docs/pricing?hl=ru.

11 Vertex AI Generative AI Service Specific Terms // Google Cloud. – URL: https://cloud.google.com/terms/service-terms.

12 Republic of Kazakhstan. Laws. On Personal Data and Its Protection: Law of the Republic of Kazakhstan No. 94-V of 21 May 2013 (as amended and supplemented as of 17.11.2025). – URL: https://adilet.zan.kz/rus/docs/Z1300000094.

13 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation). – URL: https://gdpr-info.eu/.

Авторлар туралы ақпарат // Информация об авторах // Information about the Authors

Жакупбекова Диана Бекайдаровна – студентка 3 курса по специальности «Программное обеспечение», ТОО Колледж «Astana IT University», г. Астана, Казахстан.
Жакупбекова Диана Бекайдаровна – «Бағдарламалық жасақтама» мамандығының 3-курс студенті, «Astana IT University» Колледжі ЖШС, Астана қ., Қазақстан.
Diana Zhakupbekova – 3rd-year student in "Software Engineering" specialty, LLP College «Astana IT University», Astana, Kazakhstan.
e-mail: studingzhakdi@proton.me
ORCID iD: 0009-0006-4008-0174

Мартынцов Николай Викторович – Магистр, Старший преподаватель, ТОО "Astana IT University", г. Астана, Республика Казахстан
Мартынцов Николай Викторович – Магистр, аға оқытушы, "Astana IT University" ЖШС, Астана қ., Қазақстан Республикасы
Martyntsov Nikolay Viktorovich -Master's Degree, Senior Lecturer, Astana it University LLP, Astana, Republic of Kazakhstan
e-mail: Nikolay.Martyntsov@astanait.edu.kz
ORCID iD: 0009-0000-4580-691