*IFSCC 2025 full paper (**IFSCC2025-1305**)*

An Innovative AI-Based Pipeline for More Efficient and Sustainable Botanical Extracts

Coralie Ebert[1], Ophir Kantor[1], Marie-Jose Epaule Chauvin[2], Hilla Ben-Hamo Arbel[1], Bat-Chen Tamim-Yecheskel[1], Lysiane Tang[3], Jean-Luc Ansel[4]

[1] MeNow Ltd., Israel;

[2] Ornatum Cosmetologie, France

[3] CHESE Cluster, Madagascar

[2] Cosmetopee Consulting, France

Abstract

Consumers increasingly seek natural ingredients, often perceived as more sustainable than synthetic alternatives. However, botanical extraction processes frequently generate significant waste, challenging their environmental impact.

This study presents an AI-based methodology designed to optimize plant utilization and minimize extraction waste. Structure-based classification models were used to predict the molecular composition of extracts and byproducts, while QSAR models and deep Bayesian networks evaluated potential biological benefits and mechanisms of action. For each plant part, the AI system generated multiple extraction pipelines to maximize beneficial by-products while reducing residual waste.

We applied this methodology to *Calophyllum inophyllum* (Tamanu oil), traditionally extracted by cold pressing from seeds. Our approach optimized pipelines for seeds, fruits, and leaves, achieving over 70% waste reduction and generating extracts with diverse predicted benefits, including anti-aging, anti-alopecia, and sensitive skin applications. Additionally, novel flavonoids exclusive to *Calophyllum* were identified with high predicted skincare efficacy.

This work highlights the potential of AI to revolutionize botanical extraction, enhancing sustainability, resource efficiency, and innovation in natural cosmetic ingredient development.

1. Introduction

The growing consumer demand for natural and sustainable cosmetic ingredients has driven the cosmetic industry to increasingly turn toward botanical extracts.[1-2] Natural ingredients are often perceived as safer, more environmentally friendly, and better aligned with clean beauty trends compared to synthetic alternatives.[3-5] However, the production of botanical extracts is not without challenges: traditional extraction methods often generate significant agricultural and industrial waste, raising concerns about the true sustainability of these products. As an example, to obtain 1kg of rose essential oil via steam distillation, 3000kg of flowers are needed, generating huge amounts of waste waters that are not exploited even though they contain many bioactive compounds.[6] Additionally, large volumes of by-products generated by other industries, such as the food and agricultural sectors, remain underutilized, representing a significant opportunity for sustainable innovation in cosmetic ingredient development.[7-8] Embracing such strategies aligns closely with circular economy principles, aiming to valorize waste streams, extend the life cycle of resources, and minimize environmental impact.[9]

Optimizing the use of plant resources is therefore critical to reconciling the appeal of natural ingredients with environmental responsibility. Beyond improving extraction efficiency, it is important to explore the full potential of different plant parts — including seeds, fruits, leaves, and stems — to identify novel bioactive compounds and maximize beneficial outputs. Yet, experimentally testing every possible extraction strategy across thousands of plant species and tissues is neither practical nor economically viable.

Advances in Artificial Intelligence (AI) offer a powerful solution to these challenges. By integrating structure-based classification models to predict extract compositions with QSAR modeling and deep Bayesian networks to assess potential biological benefits, AI-driven platforms can generate optimized extraction pipelines. Such approaches can identify innovative bioactives, reveal new mechanisms of action, and drastically reduce waste by promoting the use of previously overlooked plant materials.

In this study, we present an AI-based methodology that predicts optimal extraction strategies for maximizing beneficial outputs and minimizing waste generation. As a case study, we applied this platform to *Calophyllum inophyllum* (Tamanu oil). Currently, only the seeds are used in cosmetics, primarily through cold pressing to produce Tamanu oil.[10] While cold pressing is relatively low-impact, it still generates a significant amount of waste, and the potential of the fruits and leaves remains largely untapped despite strong preliminary indications of skincare benefits. For example, anti-inflammatory compounds were previously identified in the fruits.[11] Additional phytochemicals with health benefits were reported in the leaves.[12-13] Using the AI-driven platform, we demonstrated how seeds, fruits, and leaves can be better exploited to yield extracts with diverse predicted cosmetic benefits while significantly enhancing sustainability. This work illustrates the transformative potential of AI technologies to improve both environmental and scientific outcomes in natural ingredient development.

2. Materials and Methods

2.1. AI Models for Natural Products Prediction

The pipeline developed in this study incorporates three types of AI models specifically adapted for natural product analysis:

1. **Structure-based algorithms** for predicting the molecular composition according to the part of the plant and the extraction method. These models were developed using ensemble learning approaches, trained on a combination of the NPASS dataset and automatically screened literature sources. Given the presence of some molecules across multiple plant parts, a higher false-positive rate was expected.
2. **QSAR models** for predicting the biological benefits of the molecules and evaluating their relevance for various skin conditions. The detailed description and validation of these QSAR and deep Bayesian network models will be provided in a sister publication that will be published concurrently.
3. **Deep Bayesian networks** for modeling synergistic effects and mechanisms of action (MOAs) of compound combinations. The comprehensive methodology for these models will also be discussed in the sister publication.

The structure-based plant part prediction models were evaluated using 5-fold cross-validation, and their performance metrics are summarized in Table 1. Similarly, extraction method prediction models achieved F1-scores and accuracies ranging from 0.85 to 0.95, depending on the extraction method. Predictions were developed for steam distillation, CO_2 supercritical extraction, cold pressing, and solubility in ethanol, methanol, water, and glycerin.

Table 1. 5-fold cross-validation results for the structure-based prediction of part of plants

Model	FP	TP	FN	TN	Accuracy	F1-score
Seed	46	221	94	760	0.875	0.759
Flower	47	122	22	69	0.735	0.78
Root	79	443	122	962	0.875	0.815
Fruit	78	291	97	707	0.851	0.769
Leaves	96	163	49	1005	0.89	0.692

2.2. Work Pipeline

The full workflow is summarized in Figure 1. It starts by predicting the distribution of molecules across different plant parts using the plant part models.

Next, extraction models predicted which molecules could be extracted using specific methods for each plant part. Based on these predictions, the molecular composition of the initial extraction product was defined.

Residual by-products — molecules not predicted to be extracted in the first step — were then subjected to additional extraction steps. This secondary extraction was performed if more than 5% of the initial predicted bioactive compounds remained unextracted. The process could be iterated up to two additional extraction cycles depending on the remaining bioactive content.

For each extraction product generated during the process, the Skin Condition Score was predicted based on the biological activities of the extracted molecules, allowing the estimation of the potential cosmetic benefits of each extraction pipeline.

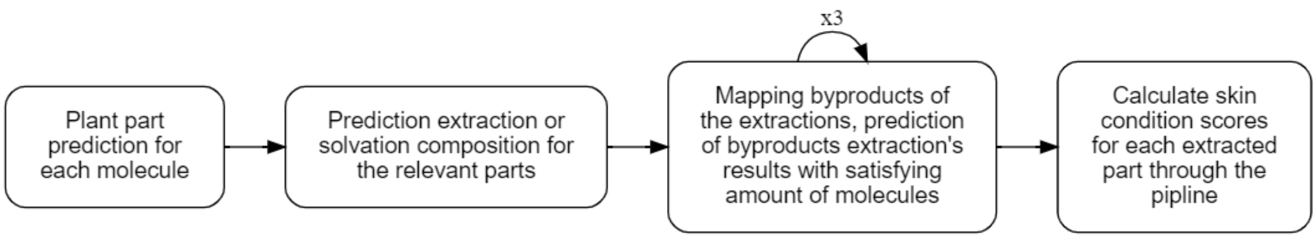

Figure 1. AI generation of extraction pipelines and computation of associated skin benefits.

2.3. Application to *Calophyllum inophyllum*

This methodology was applied to *Calophyllum inophyllum*, a plant originating from Madagascar which contains significant phytochemicals (including flavonoids, coumarins, fatty acids, and xanthones [3]) to identify new extraction pipelines for the seeds, fruits and leaves.

Based on existing databases and the literature, 335 bioactive compounds were identified in this plant. By bioactives, we mean that they have predicted activity on at least one molecular target with relevance to the skin.

3. Results

3.1. Prediction of the Molecular Composition

As mentioned previously, 335 bioactive compounds were identified in *Calophyllum inophyllum.* Of those compounds, the described algorithm predicted 188 molecules to be present in the seeds, 48 in the leaves and 117 molecules in the fruits.

Over one hundred extraction pipelines were generated according to the methodology described previously and the main skin benefits were predicted for each extract according to the molecular composition. An example of a subset of such extraction pipelines is presented for the Tamanu seeds in Figure 2.

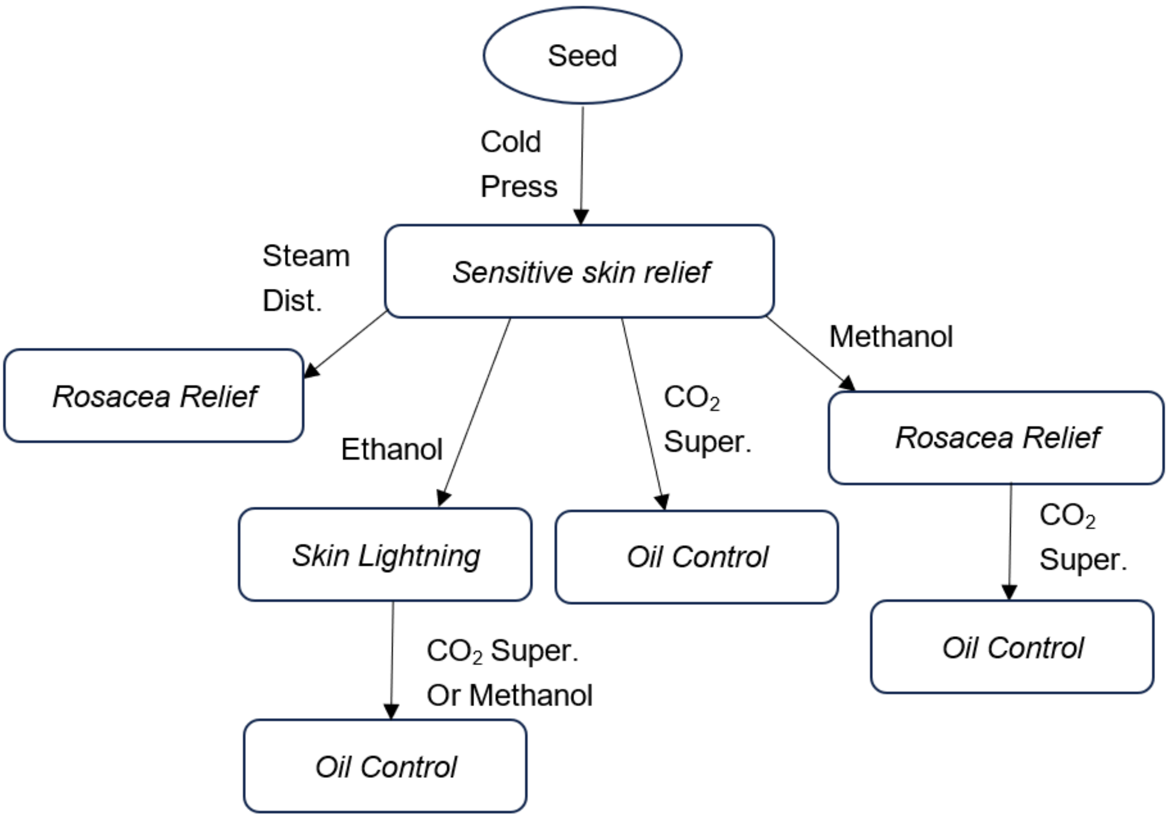

Figure 2. Extraction pipelines from *Calophyllum inophyllum*'s seeds.

3.2 Best extraction pipeline

For Tamanu oil (cold-pressed product of the *Calophyllum inophyllum* seed), high scores were predicted for photoaging protection, wound healing, anti-aging, skin rejuvenation, anti-atopic, rosacea relief, acne relief, sensitive skin relief.

Many other extracts in the seeds, leaves and fruits were identified as having high potential skin benefits for a variety of skin conditions. The seven best extraction pipelines with the top predicted skin benefits associated are presented in Table 2.

Part of the Plant	**Pipeline**	**Top Skin Conditions**
Seed	Cold-pressed (1) → Ethanol (2) → CO_2 supercritical (3)	Sensitive skin relief (1), Skin lightening (2), Oil control (3)
Seed	Cold-pressed (1) → Ethanol (2) → Methanol (3)	Sensitive skin relief (1), Skin lightening (2), Oil control (3)
Leaves	Glycerin (1) → Ethanol (2)	Skin lightening (1), Oil control (2)
Leaves	Methanol (1) → Ethanol (2)	Antiperspirant (1), Sensitive skin (2)
Fruit	CO_2 (1) → Ethanol (2)	Oil control (1), Wound healing (2)
Fruit	Methanol (1) → Ethanol (2) → CO_2 (3)	rosacea (1), skin lightening (2), oil control (3)
Fruit	Ethanol (1) → CO_2(2)	Alopecia (1), Skin rejuvenation (2)

Table 2. Top extraction pipelines according to the predicted skin benefits.

4. Discussion

4.1. Model predictions for Tamanu oil are in agreement with current knowledge.

The AI model predicted a holistic effect for Tamanu oil (cold-pressed extract of the seeds), with an effect of many skin conditions such as photoaging protection, wound healing, anti-aging, skin rejuvenation, anti-atopic, rosacea relief, acne relief, sensitive skin relief. This is in agreement in the literature, where a wide variety of skin benefits have been reported for this extract, in particular anti-inflammatory properties,[10,14-15] anti-oxidant properties, antibacterial properties, wound healing,[16] anti-atopic,[17] photoaging protection,[18] acne relief and skin rejuvenation. This concordance between predictions and experimental findings highlights the strong predictive capabilities of our AI model in accurately anticipating the broad biological effects of natural extracts.

4.2 Predictions of the molecular composition enable the use of underutilized plant parts and by-products.

While data on the molecular composition of plants exists, it is often general (covering the whole plant, as seen in databases like CMAUP [19] or LOTUS [20]) or focused on commonly used extraction methods and parts. Specific information on individual plant parts or extraction by-products is typically lacking. This limitation hampers the predictive capabilities for underutilized resources, such as the leaves and fruits of *Calophyllum inophyllum*. The methodology described here addresses this gap, offering a way to predict the bioactive potential of these less-studied parts and their by-products.

4.3. The described methodology enables the identification of high-value extraction pipelines.

By applying the described methodology to the seeds, fruits, and leaves of *Calophyllum inophyllum*, we were able to generate hundreds of potential extraction pipelines and estimate their predicted skin benefits. Among these, seven extraction strategies stood out, producing extracts with exceptionally high predicted benefits (cf Table 2). Depending on the selected extraction pipeline for each plant part, between 7 to 8 distinct extracts could be obtained, compared to only one previously available, resulting in a yield increase of over 70%.

Importantly, the resulting extracts cover a wide variety of skin conditions, with each extract showing very distinct benefit and efficacy profiles, extending far beyond the properties of the original cold-pressed seed extract. For example, the glycerin extract from the leaves was predicted to have strong anti-pigmentation properties, illustrating how targeting underutilized plant parts can diversify and expand the cosmetic applications of a single plant species. Moreover, since this methodology is not a black-box approach, it enables the identification of underlying mechanisms of action and specific bioactive compounds responsible for the predicted benefits. In the case of the glycerin extract from the leaves, two rare flavonoids, mesuein and pyranoamentoflavone, were identified as key contributors to the anti-pigmentation activity, highlighting the platform’s ability to pinpoint novel bioactive ingredients that have not been previously described.

The pipeline shows great promise for expanding the development of cosmetic ingredients, enabling the creation of multifunctional, sustainable extracts from underutilized plant parts and extraction by-products. By tapping into this previously overlooked resource, it is possible to design innovative cosmetic actives with targeted skin benefits while significantly improving sustainability profiles.

Beyond cosmetics, this approach could also have broader applications in the agro-food industry, which generates approximately 1.3 billion tons of residues annually, including skins, peels, seeds, leaves, and other inedible parts [4]. These by-products are known to be rich in secondary metabolites associated with beneficial health effects [5].

Future development will focus on integrating toxicity and odor prediction models into the scoring of extraction outputs, enabling even more complete assessment of extract viability and waste minimization potential.

5. Conclusion

Using our AI-based approach, we successfully developed optimized extraction pipelines for the different parts of the waste of Tamanu oil extraction from *Calophyllum inophyllum*, identifying multiple beneficial by-products while significantly reducing waste. By doing so, we could decrease by over 70% the amount of waste generated and generate extracts with a diverse range of predicted benefits from anti-aging to alopecia and sensitive skin. We were also able to identify highly predicted skincare benefits in the plant. This methodology demonstrates the potential of AI to revolutionize the botanical extraction industry by enhancing both sustainability and resource efficiency, as well as by identifying innovative bioactives in the extracts.

Acknowledgments

This work was supported by funding from the Israel Innovation Authority (IIA). The authors gratefully acknowledge the IIA's support, which made the development of this research possible.

References

1. Herayati, Herayati, and Elianasari Elianasari. "Natural-Based Cosmetics: Trends, Challenges, and Scientific Innovations." *Indonesian Journal of Cosmetics* 2.2 (2024): 55-81.
2. Liu, Ji-Kai. "Natural products in cosmetics." *Natural products and bioprospecting* 12.1 (2022): 40.
3. Desam, Nagarjuna Reddy, and Abdul Jabbar Al-Rajab. "The importance of natural products in cosmetics." *Bioactive natural products for pharmaceutical applications*. Cham: Springer International Publishing, 2020. 643-685.
4.
5. Simão, Sofia Alexandra Vieira, Simoni F. Rohden, and Diego Costa Pinto. "Natural claims and sustainability: The role of perceived efficacy and sensorial expectations." *Sustainable Production and Consumption* 34 (2022): 505-517.
6. Gateva, Svetla, et al. "The anti-genotoxic activity of wastewaters produced after water-steam distillation of Bulgarian Rosa damascena Mill. and Rosa alba L. essential oils." *Life* 12.3 (2022): 455.
7. Krzyżostan, Monika, Agata Wawrzyńczak, and Izabela Nowak. "Use of waste from the food industry and applications of the fermentation process to create sustainable cosmetic products: A review." *Sustainability* 16.7 (2024): 2757.
8. Lukić, M.; Ćirić, A.; Božić, D.D.; Stanković, J.A.; Medarević, Đ.; Maksimović, Z. Extracts from Wheat, Maize, and Sunflower Waste as Natural Raw Materials for Cosmetics: Value-Added Products Reaching Sustainability Goals. *Pharmaceutics* **2024**, *16*, 1182. https://doi.org/10.3390/pharmaceutics16091182
9. Barbulova, A.; Colucci, G.; Apone, F. New Trends in Cosmetics: By-Products of Plant Origin and Their Potential Use as Cosmetic Active Ingredients. Cosmetics 2015, 2, 82–92.
10. Raharivelomanana, Phila, et al. "Tamanu oil and skin active properties: from traditional to modern cosmetic uses." *Oléagineux, Corps Gras, Lipides* 5.5 (2018): D504.
11. Zakaria, Muhamad Bin, Zul Ilham, and Nur Airina Muhamad. "Anti-inflammatory activity of Calophyllum inophyllum fruits extracts." *Procedia chemistry* 13 (2014): 218-220
12. Susanto, David Febrilliant, et al. "Calophyllum inophyllum: beneficial phytochemicals, their uses, and identification." (2019): 1-18.
13. Susanto, David Febrilliant, et al. "Identification of phytochemical compounds in Calophyllum inophyllum leaves." *Asian Pacific Journal of Tropical Biomedicine* 7.9 (2017): 773-781.
14. Rodriguez, Dawn Shipley. "Tamanu oil." *Aromatherapy Journal* (2020): 41-46.

15. Marciano, Rachel, et al. "Tamanu Oil in Acne Management: Potential Anti-Inflammatory and Wound-Healing Properties for Scar Reduction." *Journal of Biomedical Science and Engineering* 17.10 (2024): 185-193.
16. Ansel, Jean-Luc, et al. "Biological activity of Polynesian Calophyllum inophyllum oil extract on human skin cells." *Planta medica* 82.11/12 (2016): 961-966.
17. Pribowo, Amadeus, et al. "Potential of Tamanu (Calophyllum inophyllum) Oil for Atopic Dermatitis Treatment." *Evidence-Based Complementary and Alternative Medicine* 2021.1 (2021): 6332867.
18. Gunawan, Yoselyn1; Pangkahila, Alex2; Darwinata, Agus Eka3. Topical administration of Tamanu Oil (Calophyllum inophyllum) inhibits the increase of matrix metalloproteinase-1 (MMP-1) expression and decreases of collagen dermis amount in male Wistar rats exposed to ultraviolet B. Neurologico Spinale Medico Chirurgico 4(3):p 114-118, November 2021. | DOI: 10.36444/nsmc.v4i3.186
19. Xian Zeng, Peng Zhang, Yali Wang, Chu Qin, Shangying Chen, Weidong He, Lin Tao, Ying Tan, Dan Gao, Bohua Wang, Zhe Chen, Weiping Chen, Yu Yang Jiang, Yu Zong Chen, CMAUP: a database of collective molecular activities of useful plants, Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D1118–D1127
20. Rutz, Adriano, et al. "The LOTUS initiative for open natural products research: knowledge management through Wikidata." *BioRxiv* (2021): 2021-02.