information

MDPI

*Article*

# SCRATCH-AI: A Tool to Predict Honey Wound Healing Properties

**Simona Martinotti, Stefania Montani, Elia Ranzato and Manuel Striani ***

DiSIT—Department of Science, Technology and Innovation, Università del Piemonte Orientale, Viale Teresa Michel 11, 15121 Alessandria, Italy; simona.martinotti@uniupo.it (S.M.); stefania.montani@uniupo.it (S.M.); elia.ranzato@uniupo.it (E.R.)

* Correspondence: manuel.striani@uniupo.it

**Abstract**

In this work, we propose SCRATCH-AI, a tool which relies on interpretable machine learning (ML) methods (namely, Bayesian networks and decision trees) to classify honey samples into wound healing categories. Classification explores the impact of botanical origins (i.e., honey type) and key chemical–biological characteristics such as antioxidant activity on healing, assessed through wound recovery metrics. The obtained classification performance results are very encouraging. Moreover, the models provide non-trivial insights about the causal dependencies of some specific honey features on wound healing properties and show the effect of different honey types (other than the well known Manuka) on cicatrization. The tool is inherently interpretable (due to the chosen ML techniques) and made user-friendly by a carefully designed graphical interface. We believe that the information provided by our tool will allow biologists and clinicians to better utilize honey, with the ultimate goal of leveraging honey capability to accelerate healing and reduce infection risks in clinical practice.

**Keywords:** machine learning; honey; wound healing; explainable AI



Academic Editor: Gholamreza Anbarjafari (Shahab)

Received: 15 July 2025
Revised: 5 September 2025
Accepted: 18 September 2025
Published: 24 September 2025

**Citation:** Martinotti, S.; Montani, S.; Ranzato, E.; Striani, M. SCRATCH-AI: A Tool to Predict Honey Wound Healing Properties. *Information* **2025**, *16*, 827. https://doi.org/10.3390/info16100827



## 1. Introduction

Honey has been used traditionally for centuries to cure different kinds of diseases [1–3]; at present, both the general public and traditional medical practitioners are beginning to recognize honey as a dependable and effective therapeutic agent [4–6]. The remarkable properties of honey are complex since many distinct bioactive compounds are involved. Antibacterial properties are its most well-known benefit, but other health-promoting characteristics of honey have been reported in recent years, such as wound healing abilities [7–9]. The primary skin cell types, keratinocytes and fibroblasts, and endothelial cells, can all benefit from honey's potential to positively affect their behavior and phenotype. Therefore, honey is becoming a very appealing option for tissue regeneration and wound repair techniques [4–6].

Each plant species has a distinct metabolic profile, and the types and quantities of plant molecules found in honey might vary based on where they come from; therefore, the honey type summarizes a set of biological characteristics that make it unique. The honey content in terms of polyphenols and other positive molecules varies significantly depending on its botanical origin as well as on other parameters (such as climate conditions, geographical area, etc.).

In this work, we aimed at analyzing the effect of such features on wound healing, with particular attention to three local honey types, whose effect was still under-investigated.

To this end, we have implemented **SCRATCH-AI** (*Smart Classification and Recovery Analysis Tool for Tissue Cicatrization and Honey-based Artificial Intelligence*), a tool which adopts machine learning (ML) techniques to classify honey samples into four healing-favoring classes, assessed through wound recovery metrics. The tool relies on Bayesian Networks and decision trees, two approaches which are recognized for their interpretability [10,11]: unlike other ML techniques, they do not work as black boxes but provide an output where the impact of the different input features on classification results is explicitly shown to the final user. This characteristic makes the tool immediately useful for biologists and clinicians, who, supported by a user-friendly interface, can gather information to create personalized therapeutic plans exploiting honey properties.

Through the tool, we have obtained very encouraging classification results on the available dataset. Moreover, the models have provided non-trivial insights about the causal dependencies of some specific honey features on wound healing properties.

*Novelty and Contributions Summary*

- SCRATCH-AI is a graphical tool that unifies Bayesian networks and decision trees to classify honey samples into four wound-healing categories from biochemical markers and botanical origin. Unlike prior software focused on scratch-assay measurement from microscopy images (PyScratch v. 2 [12]), SCRATCH-AI provides two powerful but easily interpretable ML instruments to reveal how biological features interact and impact the four categories.
- The graphical user interface (GUI) supports real-time validation and expert feature weighting, aligning with laboratory workflows for non-programmers users.
- The work emphasizes understudied local honeys (Piedmont (Italy): chestnut, coriander, sunflower) and reports 10-fold cross-validated results with per-class metrics and confusion matrices, yielding actionable biological insights alongside accurate predictions.

## 2. Related Work

Natural ethnopharmacological products like honey are used all over the world and in the rapidly evolving fields of science and technology. This classic nutrient has been suggested for a number of novel applications due to its rich chemical composition. The fundamental mechanisms behind honey's many health advantages depend on its distinct chemical environment and its nutritional wealth of polyphenols, flavonoids, bioenzymes, amino acids, and $H_2O_2$.

Honey wound-healing properties at the cellular and molecular level have been demonstrated in [13,14], where the authors show that its beneficial effects on skin cells are directly tied to its $H_2O_2$ content. Aquaporin-3 allows $H_2O_2$ extracellularly generated to pass through plasma membranes, causing an increase in $Ca^{2+}$ in the cytosol and ultimately leading to wound closure cell signalling [13,14].

The correlation between the chemical and physical characteristics of honeys (such as TPC—total phenolic content, TFC—total flavonoid content, and $H_2O_2$ content) and their biological activity related to wound healing has also been extensively proven. In this regard, our previous work [15,16] was the first to establish that the parameters we are using in this article are valid and applicable for evaluating the efficacy of different honeys in promoting wound closure.

Certain types of honey have been demonstrated to be effective in healing wounds, but the majority of recent studies (see, e.g., [17]) have concentrated on Manuka honey, a type made in New Zealand from the nectar of the Leptospermum Scopartum shrub. Indeed,

for clinical usage, a number of businesses gather, collect, filter, and sterilize Manuka honey. While the filtration eliminates wax, grit, and pollen particles from the honey to lessen the possibility of an allergic reaction, the gathering and pooling of the honey helps limit batch-to-batch variability between hive locations and seasons.

Manuka honey is expensive and not largely available. Therefore, there is an increasing interest in the characterization of the biological properties of other honey types, which could be able to stimulate wound healing, and which might be available at lower prices; the present work is located within this research area.

There is undoubtedly a need for simpler instruments to test honeys and quickly comprehend their potential, as the biological tests that are currently available on the qualities that make a honey a valuable and intriguing product for wound repair are difficult and expensive to carry out. Leveraging the power of machine learning tools to extract information from the available data is therefore an attracting research direction.

A few works have adopted ML techniques to investigate honey properties. The study in [18], for instance, has classified honey samples as either "natural" or "processed" based on eleven legally recognized quality-control parameters, using different techniques, including linear discriminant analysis (LDA) and k-nearest neighbors (KNN) classification. A key finding was that only two features—total acidity and diastase—were sufficient to achieve a near-complete separation, suggesting that more variables can introduce noise.

More similarly to what we propose, the work in [12] has integrated honey chemical fingerprinting with in vitro assays and an interpretable ML framework (called PyScratch) to evaluate wound-healing capabilities; the tool automates quantitative assessments of scratch assays, showcasing how dedicated software can streamline wound-closure measurements. Our work proceeds in the same direction, concentrating on the role of honey type on wound healing capabilities. However, our approach is significantly different. Indeed, PyScratch [12] is a tool for automatically inferring cell migration from microscopy images, to perform quantitative analysis of "in vitro" scratch assays in a fast way. On the other hand, SCRATCH-AI is an ML classification tool to predict wound healing, which provides the final user, supported by an easy GUI, with explainable results, and can reveal interesting interactions among biological markers. Table 1 summarizes the main differences between the two systems.

**Table 1.** Comparison between PyScratch and SCRATCH-AI tools.

| **Dimension** | **PyScratch ([12])** | **SCRATCH-AI (This Work)** |
|---|---|---|
| Primary objective | Automate scratch-closure measurement from images | Classify honey samples into wound-healing categories |
| Input modality | Microscopy images of scratch assays | Biochemical markers (DPPH, TFC, TPC, ROS) + HONEY TYPE |
| Core methods | Image processing by using OpenCV python library | Bayesian Networks (BN) + Decision Trees (DT) in one pipeline |
| Main outputs | Quantitative closure metrics (e.g., % wound area over time) | Predicted class (4 levels) + BN graph of dependencies + DT rules |

**Table 1.** *Cont.*

| Dimension | PyScratch ([12]) | SCRATCH-AI (This Work) |
|---|---|---|
| GUI/usability | Interface tailored to image analysis workflow | Desktop GUI for non-programmers: real-time input checks, "show model" toggle, expert feature-weighting (0–1) |
| Domain emphasis | General scratch-assay image pipelines | Under-studied local honeys (e.g., chestnut, coriander, sunflower; beyond Manuka) |
| Evaluation focus | Accuracy/quality of closure measurement | Stratified 10-fold CV with per-class metrics, confusion matrices |
| Limitation addressed | Manual/subjectiveimage measurement | White-box ML: provides transparent dependencies and explainable information |

## 3. Methods

### *3.1. Data Collection*

**Chemicals and reagents.** All reagents, if not otherwise specified, were purchased from Merck (Milan, Italy).

**Cell Culture.** L-glutamine (200 mM), 10% fetal bovine serum (FBS, Euroclone, Pero, Italy), and antibiotics (pen/strep) were added to the DMEM (high glucose, 4.5 g/L) used to cultivate the human spontaneously immortalized HaCaT keratinocytes (RRID:CVCL_0038) at 37 °C in a 5% $CO_2$ humidified atmosphere.

**Honey Type**. The honeys were obtained from "Davide e le sue Api" (https://www.ilgolosario.it/it/miele-davide-taverna, (accessed on 1 June 2025)), a local beekeeper, and collected in the Piedmont area (Northwest Italy). We utilized three different types of honey, i.e., chestnut, coriander, and sunflower. Honey samples were sourced during the summer 2024 harvest. All honey varieties (e.g., chestnut, sunflower, coriander) were collected from a distinct honey batch, defined as all honey from a specific floral source collected from a single apiary during the same harvest period. Upon collection, all samples were immediately transferred to opaque, airtight containers to protect them from light exposure. They were subsequently stored at 4 °C until analysis to preserve their physicochemical properties and prevent the degradation of bioactive compounds.

**Total Phenolic Content (TPC) assay.** We used the colorimetric in vitro Folin–Ciocalteu method to determine the amount of phenol present (TPC, total phenolic content) in the honey samples under investigation [7–9]. 250 µL of Folin Ciocalteu (10% $v/v$) was mixed with 100 µL of honey solution (1 gr in 3 mL of $H_2O$). 100 µL of 7.5% ($w/v$) $Na_2CO_3$ was added after 5 min, and the mixture was incubated at 50 °C for 15 min. After that, samples were washed with 2.5 milliliters of water. With a plate reader (Infinite 200 Pro, Tecan), the absorbance was measured at 620 nm. The results were represented in milligrams of gallic acid equivalents ($mgGAE$ 100 $g^{-1}$), and the standard curve was defined by known quantities of gallic acid.

**Total Flavonoid Content (TFC) assay.** The total flavonoid content (TFC) of the honey samples was determined using a spectrophotometric method [7–9] based on the for-

mation of a flavonoid-aluminum complex. Briefly, a 1 mL aliquot of honey solution (0.5 g/mL) was combined with 1 mL of a 2% aluminum chloride ($AlCl_3$) solution. The mixture was incubated at 25 °C for 10 min. The absorbance was then measured at 415 nm using a plate reader (Infinite 200 Pro, Tecan). A calibration curve was established using quercetin as the standard, and the results were expressed as milligrams of quercetin equivalent ($mgQE$) per 0.5 g of honey.

**DPPH assay.** The antioxidant capacity of honeys was evaluated by the free radical scavenging ability of 2,2-Diphenyl-1-picrylhydrazyl (DPPH). The assay was performed by adding 1 mM DPPH solution to honeys (1 gr in 3 mL of $H_2O$) followed by a 10-min incubation. The absorbance was determined at 512 nm in a plate reader (Infinite 200 Pro, Tecan). The radical scavenging activity was compared to water.

**Reactive oxygen species (ROS) content.** As previously reported by [13,14], xylenol orange, a colorimetric test, was used to measure the amount of $H_2O_2$ produced by honeys. In short, a solution comprising 25 mM ammonium ferrous sulphate in 0.25 M sulfuric acid was mixed with a 1:100 solution of 125 µM xylenol orange and 150 mM sorbitol to create a working solution. Next, 20 µL of the honey sample (1 gr in 3 mL $H_2O$) was combined with 200 µL of the working solution. The absorbance was measured using a 96-well microplate reader (Infinite 200 Pro, Tecan).

**SCRATCH-score wound healing assay.** A popular and easy "in vitro" method for examining the migration and proliferation of two-dimensional cell cultures is SCRATCH-score wound healing assay. It is based on the finding that when a confluent cell monolayer is exposed to an artificial gap, or scratch, the cells on the gap's edge will multiply and migrate in the direction of the opening, gradually sealing the scratch until the monolayer is restored. In many physiological or pathological contexts, such as tissue development or wound healing, the behavior of various cellular lines can be easily and effectively studied using this "in vitro" technique, which mimics the migration and proliferation of cells "in vivo".

A sterile 0.1–10 µL pipette tip was used to scrape the confluent monolayers of *HaCaT* keratinocytes. After removing any suspended cells, cultures were re-fed with media for 24 h in presence or not of honey. After the experiment, cells were fixed in 3.7% formaldehyde for 10 min and stained with 0.1% toluidine blue for 20 min. An inverted microscope (Leica Microsystems, Wetzlar, Germany) equipped with a digital camera was used to measure the width of the wound space at the start and end of the treatment. The NIH Image J (v 1.54p, https://imagej.net/ij/) program was used to assess the digital images of wounds. The difference between the wound width at 0 and 24 h was used to calculate wound closure [19–21].

The work [19] further demonstrates the broad applicability of these same rigorous analytical methods by successfully assessing the efficacy of other natural products, such as propolis, in inducing wound healing. This showcases the versatility and reliability of our approach in evaluating a range of complex natural substances for their therapeutic potential in dermatological applications.

Figure 1 shows a SCRATCH-score assay. In particular, the top panel shows a confluent monolayer of cells with a scratch (wound) created in the center; the pink circles represent individual cells. The bottom panel shows the progressive cell migration into the scratch area over time, suggesting partial or complete wound closure as the cells proliferate and move inward.

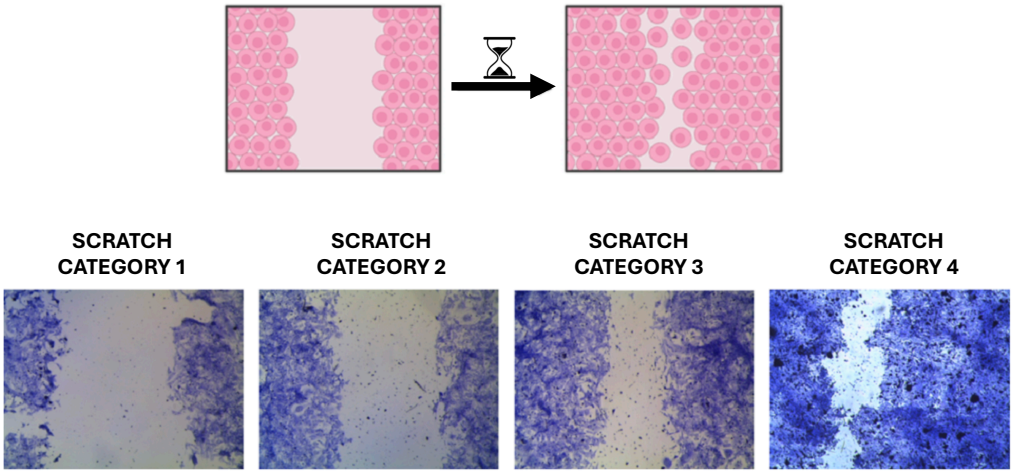



**Figure 1.** Progression of wound closure in a scratch wound healing assay (SCRATCH CATEGORY as classes).

By means of the data collection steps described above, we built a dataset of 1337 instances, each characterized by a set of biologically relevant features, which include DPPH, TFC, TPC, ROS, and the categorical variable HONEY TYPE, which represents the botanical origin of each sample. The dataset has no missing values, as the data were derived from careful laboratory measurement.

The number of instances represents the total number of individual data points collected from all experiments. In particular, for each experimental condition, we performed a minimum of 8 independent biological replicates. A biological replicate represents a completely independent experiment, from sample preparation to final measurement. Within each biological replicate, we also performed a number of technical replicates to minimize measurement variability. The number of technical replicates varied depending on the assay.

The output class SCRATCH-CATEGORY to be predicted is a discretization of the numerical continuous feature SCRATCH-score, which reflects the regenerative skincare potential of the samples. The class values are defined as follows: SCRATCH CATEGORY = 1 corresponds to SCRATCH-score values below 7.11 (low regenerative potential/minimal cell migration), SCRATCH CATEGORY = 2 spans from 7.11 to 8.0 (moderate migration), SCRATCH CATEGORY = 3 includes scores from 8.1 to 8.7 (advanced migration), and SCRATCH CATEGORY = 4 includes values above 8.7 (high regenerative potential/near-complete or complete closure). The bottom panel of Figure 1 illustrates the four classes by means of representative microscopy images.

The dataset is well balanced across these classes, with 327 samples in Class 1, 330 in Class 2, 332 in Class 3, and 348 in Class 4.

Overall, the dataset offers a well-structured and biologically meaningful basis for training interpretable ML learning models aimed at predicting SCRATCH CATEGORY and characterizing cosmetic functionality based on natural product origin and bioactivity markers.

### *3.2. SCRATCH-AI: A Tool for Predicting Wound Healing Properties*

We have developed a specialized software tool called SCRATCH-AI, designed to support and enhance wound healing predictions, in the context of scratch assay data used in cicatrization research. The tool was developed in Python v. 3.13.3 ( https://docs.python.org/release/3.13.3/, 17 September 2025) and features a Graphical User Interface (GUI) built using the Tkinter library [22] to ensure user-friendliness. On the backend, it integrates Scikit-learn v. 1.7 (https://scikit-learn.org/, 17 September 2025) [23,24], a widely used

Python module that offers a broad range of state-of-the-art ML algorithms for supervised and unsupervised learning tasks.

SCRATCH-AI incorporates both Bayesian networks and decision trees, enabling accurate, interpretable and easy-to-use classification of honey samples into classes that reflect their regenerative potential.

Indeed, Bayesian networks and decision trees are inherently interpretable ML models [10,11], which explicitly show what input features were used to make predictions, and how they were exploited.

Moreover, a key design priority was the GUI, developed leveraging the collaboration with domain-expert biologists, in order to ensure its alignment with laboratory workflows and its easy of use also by researchers without programming experience. Through the interface, users can input measured biochemical and biological variables—including DPPH, TFC, TPC, and ROS. To further optimize classification performance, each input variable is paired with a dropdown menu that allows users to assign a normalized weight within the [0–1] range, thereby leveraging domain expertise to emphasize the most biologically relevant features. A default value of 1 is provided for all weights, but experts are encouraged to edit the values, as a form of knowledge elicitation.

Each variable is entered into a dedicated field, and contextual tooltips guide users in formatting and in respecting value ranges.

Figure 2 illustrates the GUI layout. The GUI includes real-time validation, warning users when values fall outside biologically meaningful limits (error messages are shown in red). In Figure 2a, for example, the user is warned that DPPH, TFC, and TPC must be between 0 and 1, and ROS between 0 and 10. The honey type can also be selected from a dropdown menu.

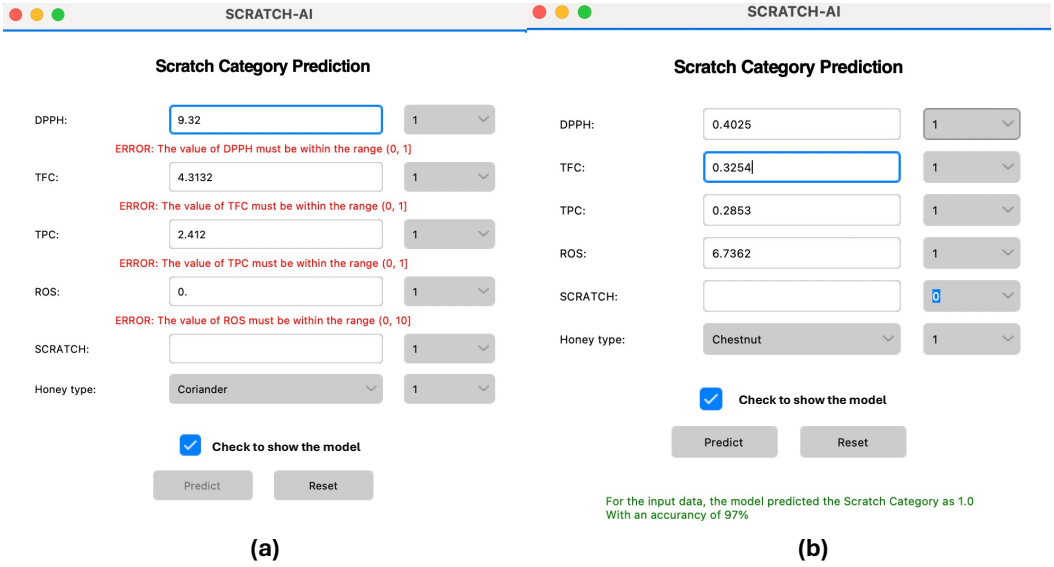



**Figure 2.** An example of **SCRATCH-AI GUI**. The left panel (**a**) illustrates input errors identified by data quality checking, while the right panel (**b**) shows a valid input scenario and prediction output with an accuracy value of 97% and a predicted SCRATCH CATEGORY of 1.

Overall, SCRATCH-AI is designed to facilitate transparent, accessible, and scientifically grounded predictions, bridging experimental biology with interpretable and explainable ML models.

## 4. Results

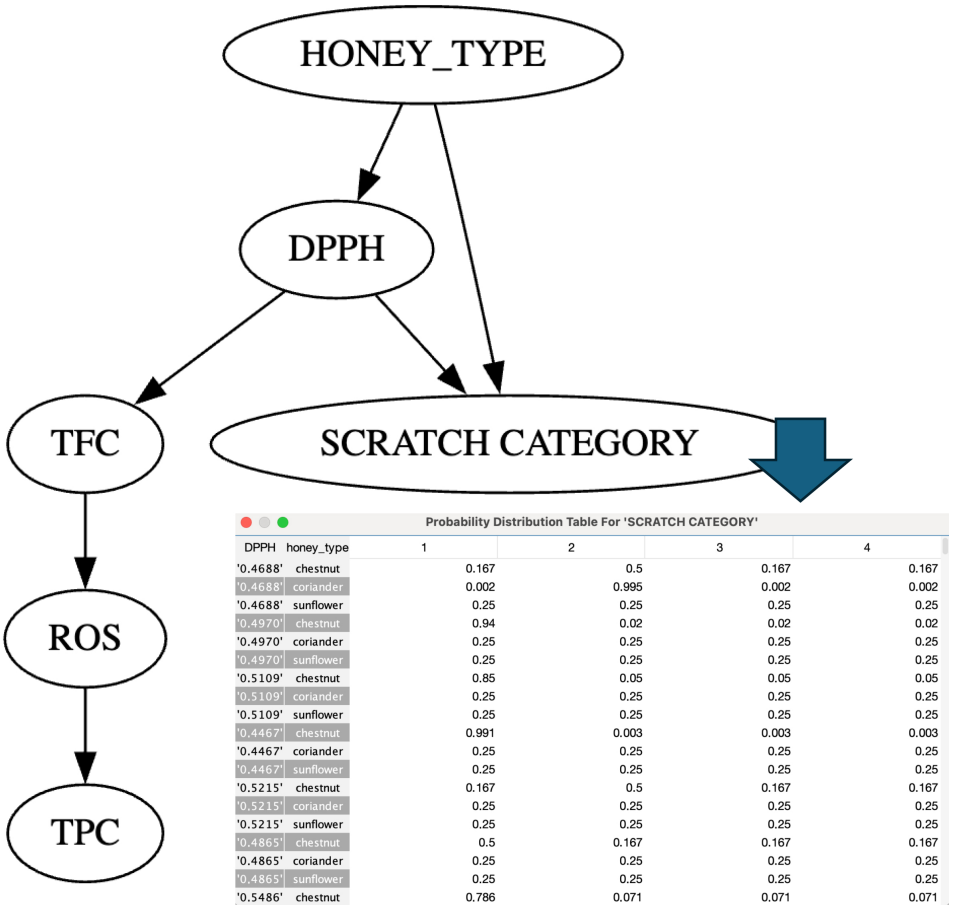
As already mentioned, our tool incorporates the possibility of adopting both Bayesian networks (BN) and decision trees (DT), to make interpretable prediction. In the following sections, we will first illustrate the BN model we obtained by working on our dataset, and then the DT one, along with the classification performance. Finally, we will show the results of a first validation study on the GUI usability.

### *4.1. A BN Model for SCRATCH CATEGORY Classification*

As is well known, a Bayesian network (BN) [25–28] is a probabilistic model that allows the graphical representation of the statistical dependence relationships between the variables of a system. Bayesian classifiers are based on a BN model, which describes the probabilistic interrelationships between features and classes.

During the training phase, the structure and probability distribution of the network are calculated in order to maximize the relationships between the data and the corresponding classes. Then, given a new item, the classifier applies the value of its features as observations in the network and then assigns the resulting class with the highest probability. In our application, we have discretized continuous variables and adopted a hill climbing structure learning algorithm.

The use of a BN allowed us to recognize which variables statistically influence others when predicting the class (outcome SCRATCH CATEGORY), as graphically represented in Figure 3, showing some non-trivial findings that will be discussed in Section 5.




Probability Distribution Table For 'SCRATCH CATEGORY'

| DPPH | honey_type | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| '0.4688' | chestnut | 0.167 | 0.5 | 0.167 | 0.167 |
| '0.4688' | coriander | 0.002 | 0.995 | 0.002 | 0.002 |
| '0.4688' | sunflower | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.4970' | chestnut | 0.94 | 0.02 | 0.02 | 0.02 |
| '0.4970' | coriander | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.4970' | sunflower | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.5109' | chestnut | 0.85 | 0.05 | 0.05 | 0.05 |
| '0.5109' | coriander | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.5109' | sunflower | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.4467' | chestnut | 0.991 | 0.003 | 0.003 | 0.003 |
| '0.4467' | coriander | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.4467' | sunflower | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.5215' | chestnut | 0.167 | 0.5 | 0.167 | 0.167 |
| '0.5215' | coriander | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.5215' | sunflower | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.4865' | chestnut | 0.5 | 0.167 | 0.167 | 0.167 |
| '0.4865' | coriander | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.4865' | sunflower | 0.25 | 0.25 | 0.25 | 0.25 |
| '0.5486' | chestnut | 0.786 | 0.071 | 0.071 | 0.071 |

**Figure 3.** Bayesian network obtained by using SCRATCH CATEGORY as a class. Conditional probability tables can be shown for every node, on demand, specifying the impact of parent node values over child node values. As an example, below the "SCRATCH CATEGORY" node, we have provided an excerpt of its conditional probability table (CPT).

The model was trained using 10-fold cross-validation. Below, we present the average cross validation results, while the model in Figure 3 refers to the best fold.

The BN classifier demonstrated strong performance, correctly classifying 1270 out of 1337 instances, yielding an overall *accuracy* of 94.99% and an error rate of 5.01%. It achieved a *K-Statistic* of 0.9332, indicating a high level of agreement beyond chance. Error metrics were notably low, with a mean absolute error ($MAE$) of 0.0344 and a root mean squared error ($RMSE$) of 0.1275. Additionally, the relative absolute error was 9.16% and the root relative squared error was 29.44%, further supporting the model's precision.

As detailed in Table 2 and in the confusion matrix in Table 3, the classifier performed consistently well across all classes. Class 1 achieved a true positive (TP) rate of 0.963 with a precision of 0.900 and an *F-measure* of 0.931, alongside an *ROC* area of 0.997. Class 2 showed exceptional results with a TP rate of 0.976, precision of 0.970, and nearly perfect ROC and PRC Areas (0.998 and 0.996, respectively). Class 3 exhibited a slightly lower recall (TP rate = 0.895), but maintained high precision (0.946), an F-measure of 0.920, and a Matthews correlation coefficient ($MCC$) of 0.894. Class 4 reached a TP rate of 0.966, precision of 0.985, and an *F-measure* of 0.975, along with the highest $MCC$ of 0.967.

**Table 2.** Results obtained by the Bayesian network.

| Class | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Accuracy |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.963 | 0.035 | 0.900 | 0.963 | 0.931 | 0.908 | 0.997 | 0.991 | |
| 2 | 0.976 | 0.010 | 0.970 | 0.976 | 0.973 | 0.964 | 0.998 | 0.996 | |
| 3 | 0.895 | 0.017 | 0.946 | 0.895 | 0.920 | 0.894 | 0.994 | 0.986 | |
| 4 | 0.966 | 0.005 | 0.985 | 0.966 | 0.975 | 0.967 | 0.997 | 0.991 | |
| **Wt. Avg.** | 0.950 | 0.016 | 0.951 | 0.950 | 0.950 | 0.934 | 0.997 | 0.991 | **0.9498** |

The weighted average metrics further emphasize the model's balanced predictive capability, with a TP rate of 0.950, precision of 0.951, *F-measure* of 0.950, and an $MCC$ of 0.934. The *ROC* and *PRC* area scores, consistently high across all classes (0.997 and 0.991 respectively), confirm the BN classifier's robustness and strong discriminative power. These results underscore the BN model's effectiveness in handling multi-class classification tasks with high reliability and precision.

The confusion matrix in Table 3 offers a detailed view of the model's performance across individual classes. Most instances are accurately classified, indicating strong predictive capability. For Class 1, the model shows high accuracy with only 12 instances misclassified as Class 3. Class 2 is also well-predicted, with minimal confusion—just 3 samples misclassified as Class 1, and a few others misassigned to Classes 3 and 4. Class 3 exhibits slightly more dispersion, with 21 samples misclassified as Class 1 and 10 as Class 2, possibly due to feature similarities across categories. Class 4 demonstrates robust classification performance, with only minor misclassification into Class 1 and Class 3. Overall, the model demonstrates strong class-wise discrimination and resilience against inter-class overlap. Despite these few misclassifications, the model maintains overall predictive consistency and robustness.

**Table 3.** Confusion matrix obtained by the Bayesian network. The ↓ indicates the actual class, while ← it indicates the predicted class.

| **Actual ↓** | $a = 1$ | $b = 2$ | $c = 3$ | $d = 4$ | **← Predicted** |
|---|---|---|---|---|---|
| $a = 1$ | 315 | 0 | 12 | 0 | |
| $b = 2$ | 3 | 322 | 4 | 1 | |
| $c = 3$ | 21 | 10 | 297 | 4 | |
| $d = 4$ | 11 | 0 | 1 | 366 | |

### *4.2. A DT Model for SCRATCH CATEGORY Classification*

A decision tree (DT) [29] is a predictive model, where each internal node represents a feature, an arc towards a child node represents a possible value for that feature, and a leaf represents the predicted class on the basis of the feature values on the branch. Classifiers based on DTs thus exploit a model that allows the class to be identified by traversing the tree from its root to one of the leaves, based on the values of the observed features. These classifiers are able to build the tree structure during the training phase, offering a model to be used for inferring the class of new objects provided as queries. In this study, we relied on the C4.5 algorithm [30] with entropy-based information gain ratio as the measure of heterogeneity for splitting nodes.

We trained a DT to predict the SCRATCH CATEGORY outcome (Figure 4). The model was trained using 10-fold cross-validation. Below, we present the average cross-validation results, while the model in Figure 4 refers to the best fold.

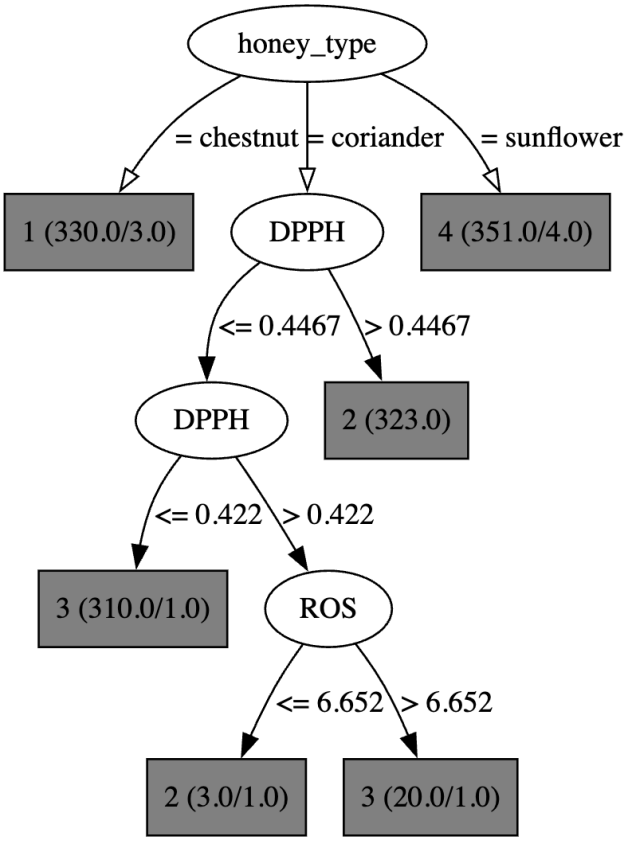



**Figure 4.** Decision tree obtained by using SCRATCH CATEGORY as a class. Gray nodes represent the class and also specify the number of items that satisfy the decision indicated on the incoming edge and that belong to the class at hand. For instance in the first gray node (upper left corner), 330 items exhibit chestnut as a honey type, and only three of them are not classified in SCRATCH CATEGORY 1.

The DT confirms some of the findings already identified by the BN in Section 4.1 and provides further insights, as discussed in Section 5.

The DT model correctly classified 1266 out of 1337 instances (*accuracy* = 94.69%) and achieved a high *K-Statistic* of 0.9291, indicating strong agreement beyond chance. Its error metrics, including a mean absolute error of 0.0276 and a root mean squared error of 0.1241, are notably low, underscoring the model's precision. Likewise, the relative absolute error of 7.3536% and the root relative squared error of 28.6651% suggest robust performance

compared to a baseline. Overall, these metrics indicate a highly reliable and effective classification performance.

Results are shown in Table 4.

**Table 4.** Results obtained by the decision tree.

| Class | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Accuracy |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.917 | 0.001 | 0.997 | 0.917 | 0.955 | 0.943 | 0.997 | 0.988 | |
| 2 | 0.976 | 0.002 | 0.994 | 0.976 | 0.985 | 0.980 | 0.992 | 0.983 | |
| 3 | 0.907 | 0.003 | 0.990 | 0.907 | 0.947 | 0.931 | 0.992 | 0.977 | |
| 4 | 0.986 | 0.066 | 0.841 | 0.986 | 0.907 | 0.877 | 0.990 | 0.972 | |
| **Wt.Avg.** | 0.947 | 0.019 | 0.954 | 0.947 | 0.948 | 0.932 | 0.993 | 0.980 | **0.9469** |

Notably, Class 4 shows the highest recall (TP rate: 0.986), albeit with slightly lower precision (0.841), indicating it captures nearly all true instances but admits some false positives from other categories. By contrast, Classes 1, 2, and 3 exhibit excellent precision (0.997, 0.994, and 0.990, respectively), reflecting a very low false-positive rate.

*MCC* remains consistently high (0.943 for Class 1, 0.980 for Class 2, and 0.931 for Class 3), though it decreases somewhat for class 4 (0.877), reflecting a trade-off between precision and recall. The *ROC* area values exceed 0.99 for all classes, demonstrating outstanding discriminatory capability, and the *PRC* area remains above 0.97, underscoring the model's high reliability.

The confusion matrix shown in Table 5 illustrates in detail the performance of the DT classifier. Overall, the classifier demonstrates strong performance, correctly classifying the majority of instances across all categories. Specifically, class 2 and Class 4 exhibit excellent classification results, with 322 and 343 correctly classified instances, respectively. Minor misclassifications occur, notably with Class 1 and Class 3, each showing some confusion predominantly with class 4 (27 and 31 instances, respectively). This pattern suggests that certain biochemical profiles associated with Class 1 and 3 overlap with Class 4, potentially due to similarities in flavonoid content and oxidative stress markers among these samples. Nonetheless, the decision tree classifier maintains robust predictive accuracy, effectively distinguishing the majority of samples across SCRATCH CATEGORIES and indicating its utility for practical classification tasks in evaluating skincare properties based on biochemical characteristics.

**Table 5.** Confusion matrix obtained by the decision tree. The ↓ indicates the actual class, while ← it indicates the predicted class.

| **Actual ↓** | $a = 1$ | $b = 2$ | $c = 3$ | $d = 4$ | **← Predicted** |
|---|---|---|---|---|---|
| $a = 1$ | 300 | 0 | 0 | 27 | |
| $b = 2$ | 1 | 322 | 0 | 7 | |
| $c = 3$ | 0 | 0 | 301 | 31 | |
| $d = 4$ | 0 | 2 | 3 | 343 | |

### *4.3. GUI Evaluation*

To evaluate the usability of our SCRATCH-AI GUI, we asked five biologists to use the tool and then complete a Likert-scale (1–5) usability questionnaire composed by five items, adapted for the SCRATCH-AI tool from [31]. Below, the structure of the questionnaire is reported (Figure 5).

# SCRATCH-AI Usability Questionnaire

**Instructions.** For each statement below, select *one* response.

**Participant ID:** _____ **Date:** _____ **Role:** _____

**Response scale**
1 = Strongly disagree; 2 = Disagree; 3 = Neither; 4 = Agree; 5 = Strongly agree

| Item | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. I found the SCRATCH-AI interface easy to use. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. I think the information shown (layout, labels, steps) is more complex than necessary. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. I felt confident interpreting the BN/DT explanations that justify a SCRATCH CATEGORY prediction. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. I needed to learn a lot before I could be productive with this interface. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. I would use this tool regularly in my laboratory/clinical workflow. | ☐ | ☐ | ☐ | ☐ | ☐ |

**Figure 5.** SCRATCH-AI usability questionnaire.

Note that we reversed the scores in questions 2 and 4, which are negative statements.

By means of this evaluation study, we obtained an average score of 4/5, indicating the high usability of SCRATCH-AI GUI.

## 5. Discussion

The BN presented in Section 4.1 identified a direct influence of the honey type on SCRATCH CATEGORY, which is indeed a hypothesis we wished to confirm, since, as observed in Section 2, the effectiveness of honey types other than Manuka on healing wounds are still under-investigated.

Moreover, the BN analysis revealed several other intriguing conditional dependencies between the phytochemical properties of honey and its biological effects. These relationships, while statistically derived, are well-supported by established biological and chemical principles. For instance, the network identified a strong conditional dependency between TFC and ROS. This finding is biologically plausible and consistent with the role of flavonoids as potent antioxidants. Flavonoids exert their antioxidant effects primarily by scavenging free radicals, chelating metal ions, and inhibiting pro-oxidant enzymes, thereby directly mitigating the formation and activity of ROS. The association identified by the network therefore represents a functional link between the antioxidant capacity of honey and its ability to modulate oxidative stress. Similarly, the network's finding of a link between DPPH radical scavenging activity and healing potential is directly aligned with the current understanding of wound biology. The DPPH assay is a widely accepted measure of antioxidant capacity. In the context of wound healing, oxidative stress plays a crucial role in perpetuating chronic inflammation and delaying tissue repair. Compounds with high antioxidant activity can protect cells from oxidative damage, modulate the inflammatory response, and create a more favorable environment for tissue regeneration. Consequently,

a higher DPPH value, indicative of greater antioxidant potential, is logically associated with a more effective "healing potential". The link between ROS and TPC suggests a potentially bivalent relationship. Phenolic compounds can either attenuate the presence of oxidizing molecules or, depending on the specific cellular context and concentration, induce ROS production. These inter-feature relationships are particularly valuable in domains such as food science or biochemistry, where molecular interactions are often non-independent and context-dependent. These examples demonstrate that the BN serves as a powerful tool for identifying and visualizing complex relationships within a dataset, providing a robust framework for further experimental validation and a deeper understanding of the biological mechanisms at play. In conclusion, the BN offers both robust classification performance (as shown in Section 4.1) and a transparent, biologically grounded structure. This makes the model a powerful tool for both predictive analytics and scientific interpretation in studies involving honey characterization, antioxidant behavior, and food classification tasks.

The DT illustrated in Section 4.2 confirms some of the findings already identified by the BN, and provides further insights. The honey type has a clear and prominent role on wound healing, with sunflower honey exhibiting the strongest positive effect: indeed, according to our current experiments, when honey type is sunflower, no other feature is actually needed to predict SCRATCH CATEGORY = 4, while chestnut, on the other hand, is not effective, as no other feature is needed to predict SCRATCH CATEGORY = 1. Coriander honey has some positive effect on wound healing too, but other features have to be considered, in order to distinguish between SCRATCH CATEGORY = 2 and SCRATCH CATEGORY = 3. More tests on additional datasets will be useful to confirm these outcomes. From a biological viewpoint, however, such findings are plausible. Sunflower honey is known to be rich in specific phenolic compounds, particularly flavonoids like quercetin and kaempferol. These antioxidants are potent free radical scavengers that can neutralize ROS. The reduction of ROS is critical in the early phases of wound healing, as it mitigates cellular damage and reduces inflammation, which are key obstacles to efficient tissue repair. Research has shown that sunflower honey often contains higher levels of the enzyme glucose oxidase, which catalyzes the conversion of glucose into gluconic acid and hydrogen peroxide ($H_2O_2$). In small, controlled concentrations, $H_2O_2$ has antimicrobial properties and can stimulate the proliferation of fibroblasts and keratinocytes, which are essential for wound closure. Like other honeys, sunflower honey has a high sugar content, creating a low water activity environment. This osmotic effect draws moisture from the wound, inhibiting microbial growth and promoting a clean healing environment. The specific phytochemicals in sunflower honey may possess anti-inflammatory properties that help modulate the inflammatory phase of wound healing. By reducing excessive inflammation, the honey can prevent chronic wound conditions and facilitate a faster transition to the proliferative phase. These factors likely work in synergy to contribute to the observed superior healing properties of sunflower honey in our model.

Consistent with the BN findings, DPPH is the second feature to be examined in order of importance. When DPPH is between 0.422 and 0.446, the distinction between the two classes further depends on ROS: in particular, the effect of a higher ROS, which may negatively affect wound healing, must be compensated by a higher DPPH antioxidant activity to reach SCRATCH CATEGORY = 3.

Overall, the results presented in Section 4.2 show that the DT classifier effectively captures the underlying biological and chemical distinctions between categories, delivering highly accurate and dependable predictions of SCRATCH CATEGORIES based on biochemical markers.

For the sake of benchmarking, we also made some classification tests resorting to a support vector machine, gradient boosting, and random forest, obtaining up to 98%

accuracy. However, these methods do not provide inherent explainable outputs. Since BN and DT performance metrics are also very high, we continue with transparent methods in SCRATCH-AI.

In regard to cross-validation, it is worth mentioning that distinct measurements from the same honey batch appear both in the training and test sets. Indeed, all samples of a specific floral type (e.g., chestnut) were harvested during a single season (2024) representing a distinct honey batch. However, the single-batch approach was chosen to minimize variability from environmental factors, seasonal changes, or beekeeping practices, allowing for a focused analysis of the specific honey types. Therefore, this choice may limit generalizability but provides strength against confounding factors. As discussed in the concluding section, in the future, we plan to extend our study, overcoming this possible limitation.

## 6. Conclusions

In this paper, we have introduced SCRATCH-AI, an interpretable ML tool able to classify honey samples into wound healing classes. By means of SCRATCH-AI, researchers can gather information to create personalized therapeutic plans exploiting honey properties, helped by a simple GUI ensuring the tool alignment with laboratory workflows.

In particular, the tool has allowed us to build a BN and a DT, that we exploited to investigate the effect of different honey types (other than the well-known Manuka) on wound healing, discovering interesting properties especially for sunflower and coriander local honey.

As a matter of fact, our research was designed as a focused, preliminary investigation to provide a detailed, in-depth characterization of the physicochemical and bioactive properties of specific, well-defined honey samples. By limiting our focus to three distinct honey types from a single, verifiable local source, we were able to minimize the confounding variables associated with beekeeping practices, environmental factors, and seasonal variations. This approach allowed for robust analysis and correlation of specific chemical profiles with their observed biological activities, establishing a foundational stepping stone for future, more expansive studies, including a wider range of honey varieties and geographical origins, ultimately leading to a more comprehensive understanding of the factors that influence honey quality and bioactivity.

In future work, we also aim to transform SCRATCH-AI from a desktop-based solution into a cloud-based web application, enabling broader accessibility and improved scalability. By containerizing the software (e.g., via Docker v. 23.0.3 and Kubernetes v. 1.34.1), the system could dynamically allocate resources based on user demand and better support large-scale collaborations among clinical and research institutions. A browser-based interface would allow researchers to securely log in from any location, leverage real-time data validation, and visualize model outputs (e.g., BN structures) in an interactive format. Integrating SCRATCH-AI with external platforms—such as laboratory information management systems (LIMS) [32] or electronic health records (EHR)—through secure APIs—would automate data ingestion, minimize manual errors, and streamline workflow.

The cloud design will follow privacy-by-design principles, such as data pseudonymization, role-based access control (RBAC) with single sign-on (SSO), and detailed audit logs.

To establish generalizability, we will work on the definition of multi-site program, including site-held-out/leave-one-center-out evaluations and targeted model adaptation (e.g., site-specific calibration layers) if needed, and fairness analyses across honey type/botanical origin and site.

If data sharing is constrained, we could enable federated learning/secure aggregation or on-prem “edge” deployments that exchange only model parameters.

The foreseen transition of SCRATCH-AI from a standalone desktop tool to a web application will also facilitate alignment with laboratory/clinical workflows: to this end, the platform will address interoperability issues and offer role-based workspaces.

A continuously updated, more diverse dataset would further enhance classification performance, enabling regular retraining of the Bayesian network and decision tree models. In addition, hosting ensemble-based approaches within the cloud environment would allow direct comparisons of multiple models in parallel, balancing the trade-offs between accuracy and interpretability.

Ultimately, a cloud-native SCRATCH-AI would advance honey-based wound research and clinical applications by facilitating collaborative data sharing, consistent model refinement, and real-time predictive analytics.

**Author Contributions:** Conceptualization, S.M. (Simona Martinotti), S.M. (Stefania Montani), E.R. and M.S.; Methodology, S.M. (Simona Martinotti), S.M. (Stefania Montani), E.R. and M.S.; Software, M.S.; Resources, S.M. (Simona Martinotti) and E.R.; Data curation, S.M. (Simona Martinotti), E.R. and M.S.; Writing—original draft, S.M. (Simona Martinotti), S.M. (Stefania Montani), E.R. and M.S.; Writing—review & editing, S.M. (Simona Martinotti), S.M. (Stefania Montani), E.R. and M.S.; Supervision, S.M. (Simona Martinotti), S.M. (Stefania Montani), E.R. and M.S. All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

**Funding:** No funding was received for this research.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Dataset available on request from the authors. The raw data supporting the conclusions of this article will be made available by the authors upon request.

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Bonsignore, G.; Martinotti, S.; Ranzato, E. Honey Bioactive Molecules: There Is a World Beyond the Sugars. *BioTech* **2024**, *13*, 47. [CrossRef]
2. Erler, S.; Moritz, R.F.A. Pharmacophagy and pharmacophory: Mechanisms of self-medication and disease prevention in the honeybee colony (*Apis mellifera*). *Apidologie* **2016**, *47*, 389–411. [CrossRef]
3. Palma-Morales, M.; Huertas, J.R.; Rodríguez-Pérez, C. A Comprehensive Review of the Effect of Honey on Human Health. *Nutrients* **2023**, *15*, 3056. [CrossRef] [PubMed]
4. Martinotti, S.; Bonsignore, G.; Ranzato, E. Applications of Beehive Products for Wound Repair and Skin Care. *Cosmetics* **2023**, *10*, 127. [CrossRef]
5. Oryan, A.; Alemzadeh, E.; Moshiri, A. Biological properties and therapeutic activities of honey in wound healing: A narrative review and meta-analysis. *J. Tissue Viability* **2016**, *25*, 98–118. [CrossRef] [PubMed]
6. Yupanqui Mieles, J.; Vyas, C.; Aslan, E.; Humphreys, G.; Diver, C.; Bartolo, P. Honey: An Advanced Antimicrobial and Wound Healing Biomaterial for Tissue Engineering Applications. *Pharmaceutics* **2022**, *14*, 1663. [CrossRef]
7. Martinotti, S.; Bucekova, M.; Majtan, J.; Ranzato, E. Honey: An Effective Regenerative Medicine Product in Wound Management. *Curr. Med. Chem.* **2019**, *26*, 5230–5240. [CrossRef] [PubMed]
8. Alam, F.; Islam, M.A.; Gan, S.H.; Khalil, M.I. Honey: A Potential Therapeutic Agent for Managing Diabetic Wounds. *Evid.-Based Complement. Altern. Med.* **2014**, *2014*, 169130. [CrossRef] [PubMed]
9. Miguel, M.; Antunes, M.; Faleiro, M. Honey as a Complementary Medicine. *Integr. Med. Insights* **2017**, *12*, 1178633717702869. [CrossRef] [PubMed]
10. Holzinger, A. From Machine Learning to Explainable AI. In Proceedings of the 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), Košice, Slovakia, 23–25 August 2018; pp. 55–66. [CrossRef]
11. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **2021**, *23*, 18. [CrossRef]

12. Garcia-Fossa, F.; Gaal, V.; de Jesus, M.B. PyScratch: An ease of use tool for analysis of scratch assays. *Comput. Methods Programs Biomed.* **2020**, *193*, 105476. [CrossRef]
13. Martinotti, S.; Laforenza, U.; Patrone, M.; Moccia, F.; Ranzato, E. Honey-Mediated Wound Healing: $H_2O_2$ Entry through AQP3 Determines Extracellular $Ca^{2+}$ Influx. *Int. J. Mol. Sci.* **2019**, *20*, 764. [CrossRef] [PubMed]
14. Liu, J.; Jin, Y.; Wei, Q.; Hu, Y.; Liu, L.; Feng, Y.; Jin, Y.; Jiang, Y. The relationship between aquaporins and skin diseases. *Eur. J. Dermatol.* **2023**, *33*, 350–359. [CrossRef] [PubMed]
15. Martinotti, S.; Bonsignore, G.; Patrone, M.; Ranzato, E. Correlation between Honey Parameters and Wound Healing Properties: The Case of Piedmont (Italy) Samples. *Curr. Pharm. Biotechnol.* **2025**, *26*, 302–311. [CrossRef] [PubMed]
16. Cianciosi, D.; Forbes-Hernández, T.Y.; Afrin, S.; Gasparrini, M.; Reboredo-Rodriguez, P.; Manna, P.P.; Zhang, J.; Bravo Lamas, L.; Martínez Flórez, S.; Agudo Toyos, P.; et al. Phenolic Compounds in Honey and Their Associated Health Benefits: A Review. *Molecules* **2018**, *23*, 2322. [CrossRef]
17. Tashkandi, H. Honey in wound healing: An updated review. *Open Life Sci.* **2021**, *16*, 1091–1100. [CrossRef]
18. López, B.; Latorre, M.; Fernández, M.; García, M.; García, S.; Herreroa, C. Chemometric classification of honeys according to their type based on quality control data. *Food Chem.* **1996**, *55*, 281–287. [CrossRef]
19. Martinotti, S.; Pellavio, G.; Laforenza, U.; Ranzato, E. Propolis Induces AQP3 Expression: A Possible Way of Action in Wound Healing. *Molecules* **2019**, *24*, 1544. [CrossRef]
20. Balko, S.; Kerr, E.; Buchel, E.; Logsetty, S.; Raouf, A. A Robust and Standardized Approach to Quantify Wound Closure Using the Scratch Assay. *Methods Protoc.* **2023**, *6*, 87. [CrossRef]
21. Walter, M.N.M.; Wright, K.T.; Fuller, H.R.; MacNeil, S.; Johnson, W.E.B. Mesenchymal stem cell-conditioned medium accelerates skin wound healing: An in vitro study of fibroblast and keratinocyte scratch assays. *Exp. Cell Res.* **2010**, *316*, 1271–1281. [CrossRef]
22. Charatan, Q.; Kans, A. Python Graphics with Tkinter. In *Programming in Two Semesters: Using Python and Java*; Springer International Publishing: Cham, Swizerland, 2022; pp. 211–254. [CrossRef]
23. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [CrossRef]
24. Kramer, O. Scikit-Learn. In *Machine Learning for Evolution Strategies*; Springer International Publishing: Cham, Swizerland, 2016; pp. 45–53. [CrossRef]
25. Lucas, P.J.F.; van der Gaag, L.C.; Abu-Hanna, A. Bayesian networks in biomedicine and health-care. *Artif. Intell. Med.* **2004**, *30*, 201–214. [CrossRef]
26. Cheng, J.; Greiner, R. Comparing Bayesian Network Classifiers. *arXiv* **2013**, arXiv:1301.6684. [CrossRef]
27. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian Network Classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [CrossRef]
28. Heckerman, D.; Geiger, D.; Chickering, D.M. Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Mach. Learn.* **1995**, *20*, 197–243. [CrossRef]
29. Mantovani, R.G.; Horvath, T.; Cerri, R.; Vanschoren, J.; de Carvalho, A.C. Hyper-Parameter Tuning of a Decision Tree Induction Algorithm. In Proceedings of the 2016 5th Brazilian Conference on Intelligent Systems (BRACIS), Recife, Brazil, 9–12 October 2016; pp. 37–42. [CrossRef]
30. Salzberg, S. Book Review: C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Mach. Learn.* **1994**, *16*, 235–240. [CrossRef]
31. Brooke, J. SUS—A quick and dirty usability scale. In *Usability Evaluation in Industry*; CRC Press: London, UK; Bristol, PA, USA, 1996; ISBN 9780748404605.
32. Skobelev, D.O.; Zaytseva, T.M.; Kozlov, A.D.; Perepelitsa, V.L.; Makarova, A.S. Laboratory information management systems in the work of the analytic laboratory. *Meas. Tech.* **2011**, *53*, 1182–1189. [CrossRef]