

## Geographical Indication Classification Based on Stingless Bee Honey Samples Using a Generative Model on Spectral Data for Mobile Device Applications

Hata Maulana<sup>1</sup>, Yohanes Aris Purwanto<sup>2\*</sup>, Diding Suhandy<sup>3</sup>, Sony Hartono Wijaya<sup>4</sup>, Heru Sukoco<sup>5</sup>

<sup>1,4,5</sup>Computer Science, School of Science, Mathematics and Informatics, Institute Pertanian Bogor University

<sup>2</sup>Mechanical and Biosystems Engineering, Faculty of Agricultural Technology, Institute Pertanian Bogor University

<sup>3</sup>Agricultural Engineering, Faculty of Agricultural, The University of Lampung

<sup>1</sup>Digital Multimedia Techniques, Informatics and Computer Engineering, Politeknik Negeri Jakarta

<sup>1,2,4,5</sup>Jl. Raya Dramaga, Bogor 16680

<sup>3</sup>Jl. Prof. Sumantri Brojonegoro No. 1, Gedong Meneng, Rajabasa, Bandar Lampung 35145

<sup>1</sup>Jl. Prof. DR. G. A. Siwabessy, Kampus UI, Depok 16425

### ABSTRACT

#### Article:

Accepted: February 13, 2026

Revised: December 01, 2025

Issued: April 30, 2026

© Maulana et al, (2026).



This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

#### \*Correspondence Address:

[arispurwanto@apps.ipb.ac.id](mailto:arispurwanto@apps.ipb.ac.id)

Stingless bee honey (SBH) is a food product with a simple and fast production process. Honey production is highly dependent on the condition and ability of the bees. Each region in Indonesia has different botanical and environmental conditions, with different entomological characteristics of bee species based on their geographical origin. This honey classification model is part of the formation of Geographical Indications (GI) based on product characteristics. A generative model is used in the pre-processing stage to produce spectrum data with the best grouping (silhouette score > 0.6). The main process discussed in this article is the application of GI classification model of four types of stingless bee honey based on cultivation location. The results of the study with a 400x745 dataset and 4 classes (Lampung, Bogor, Sukabumi, and Rangkas Bitung) showed that the classification model produced an accuracy above 95% with a precision and recall above 0.99. The SBH GI classification application has been successfully built using the scrum methodology and is cloud-based. The application displays the classification results of origin, type (entomological), feed dominance (botanical) and dominant spectrum values. The application has also been tested for feasibility based on the User Acceptance Test with results above 90%.

**Keywords :** *Stingless Bee; Honey, Spectrum; Fluorescence; Machine Learning.*

## 1. INTRODUCTION

Assessing the quality of honey is not an easy job, there are several factors that must be taken into account, such as the origin of the feed (botany) as vegetation, the origin of the region (origin) as geography, as well as the production method and storage conditions of the harvested honey [1]. In addition to these factors, observations and interviews with several beekeepers, particularly stingless bees, revealed that pure honey harvested directly from the hive is easily contaminated by improper handling and conditions. Therefore, beekeepers should immediately place the harvested honey in hygienic containers and store it at a normal temperature. The population of stingless bee traders, which is spread throughout Indonesia, has not been fully recorded. This study used a purposive method to determine the number and sample objects, namely four regions in western Indonesia.

Honey from this stingless bee species has a segmented quality and benefits, meaning only a select few consumers prefer its taste and characteristics. Its predominantly sour taste and slightly natural sweetness are due to its higher water content compared to other types of honey, such as those from stingless bees (*apis*). Stingless honey has a low sweetness level and a high water content, causing this type of honey to deteriorate more quickly [2]. Stingless bees only produce 1-5 kg of honey per year while honey bees produce an estimated 20 kg per year [3]. However, this particular market segment is expanding, with farmers reporting increasing demand. Demand for stingless bee honey (SBH) is indeed higher than for other types [4]. However, based on field observations at four sample research locations: Bogor, Lampung, Rangkas Bitung, and Sukabumi, it was found that honey production was not comparable to market demand, resulting in unstable supply. Another issue that emerged was the lack of a standardized pricing reference, preventing strategic innovations to increase production.

Production limitations and the bargaining power of these products have made SBH producers (cultivators) hesitant. Likewise, consumers (the public) desperately need to be assured of the purity of SBH products to obtain their essential benefits. Applying GI classification to SBH products is an appropriate alternative solution to gain public trust. This

application requires the use of instrument technology that is compatible with SBH products and accepted by farmers in the field.

Research in the field of spectroscopy is leading to simplified instrument use, non-destructive, and real-time operation. Industries are beginning to utilize it as an alternative technology, hoping to achieve greater profitability. This is particularly true in the food industry, which requires faster and more effective consumer trust. Based on literature reviews and preliminary observations, fluorescence spectroscopy, with its unique characteristics, provides a more sensitive and selective approach to obtaining more complete and precise characteristic information. This is because fluorescence-based spectroscopy is known to have higher sensitivity and selectivity [5].

In previous research, a fairly common method used in spectroscopy was the Partial Least Squares (PLS) method. PLS is a multivariate statistical method used to analyze the relationship between two groups of variables, each with an independent variable (predictor) and a dependent variable (response). PLS with Discriminant Analysis and Principal Component Analysis (PCA) Algorithm is used as the basis for forming classification models for supervised and unsupervised machine learning types [6]. The discriminant model suffers from overlapping data distribution or distribution of classification or clustering results. Furthermore, it lacks a pre-processing stage for using the classification model in a specific application. Probabilistic Principal Component Analysis (PPCA) focuses on spectral pattern analysis and noise reduction in spectroscopy problems [7][8][9]. PPCA is a machine learning algorithm that uses a generative model approach with a focus on data formation. The geographic indication classification model in this study uses spectral data combined with clinical trial data from stingless bee honey. This poses several challenges in the dataset preprocessing, and the use of several generative model-based machine learning algorithms was determined as a solution.

The main focus of this article is the development of the classification model using three parameters of geographical indication information from stingless bee honey. Origin, botanical, and entomological characteristics

were determined based on observations and in-depth interviews with several stingless bee breeders from the study sample locations. The results of the more complete characteristic information of this type of honey are based on the classification of the obtained machine learning model. Machine learning algorithms were used in the pre-processing stage of a honey dataset from stingless bees derived from clinical trial data and spectrum data. The dataset was then used as the basis for creating a Geographical Indication classification model to distinguish honey origins in real time within a single origin class. The honey characteristic information in this study was based on primary data collection and spectrum data measurements. The application of stingless bee honey classification is based on origin, but includes information on the dominant content characteristics of each type that differentiate the measured spectrum values. Information on the dominant food of stingless bee species based on geographic origin is classified based on the capture of dominant spectrum values with the highest main content, and so on as the uniqueness of the geographical indication of this honey product.

## 2. METHOD

This research builds on previous findings related to cloud-based application development. The application is built on three main layers as a cloud-based development architecture for implementing a geographic indication classification model for stingless bee honey [10]. Multivariate analysis approach, Vis-NIR successfully detected the contamination level of high-quality honey products using explanatory analysis, supervised analysis and unsupervised analysis [11].

Figure 1, the developed application is a prototype of the proposed new mechanisms for implementing the classification results from the obtained model. First, this study collected pure honey products from fresh harvests based on four designated research sample locations. Second, data from clinical trials and fluorescence-based spectrum capture were acquired. The use of a portable spectrophotometer was also tested in this study. Third, the final stage of mobile-based application development focused on creating geographical indications as a contribution to the

identification of this type of honey product in Indonesia.

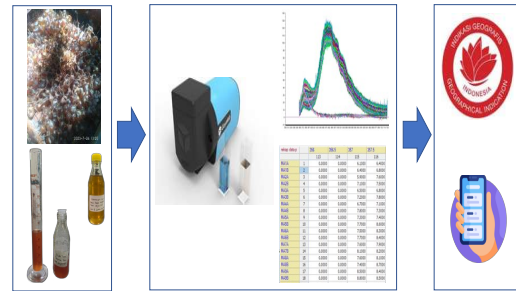


Figure 1. Application development

The spectrum data used was obtained using primary data, which directly determined the location and type of stingless bee samples used in the study. Next, the honey data obtained was subjected to clinical testing using the parameters specified in the study. The clinical test used samples from four pure honey harvest locations, each weighing 250 g.

The clinical test parameters used were six of the 12 parameters based on the Indonesian National Standard (SNI) 8664:2018: water content, diastase enzyme, HMF, reducing sugar, sucrose, and acidity [12]. Ultimately, the dataset acquisition was obtained from raw data from fluorescence-based spectrum measurements in the 357-750 nm wavelength range, while the clinical test results generated a 400x744 dataset.

Figure 2 shows the results of fresh harvest honey samples obtained from cultivators measured using a portable fluorescence spectrophotometer with a dilution ratio of 1:5, with settings on the measuring instrument with an exposure of 2000, a Gain value of 2, a Number of cycles of 10, and a Delay of 100. In addition, fresh harvest honey samples were subjected to clinical product trials in Duplo at the IPB Integrated Laboratory. Clinical trials were conducted using testing parameters based on the Indonesian National Standard (SNI). The standard was used to determine the content of honey samples from four locations: Bogor, Lampung, Sukabumi, and Rangkas Bitung. The amount of honey product samples for measurement using a portable spectrophotometer is 50-60 ml, while clinical trials require approximately 250 ml for each product sample. Fluorescence is used to identify characteristics such as proteins, polyphenol compounds, and other product reactions [5] in honey product samples from stingless bees as research objects.

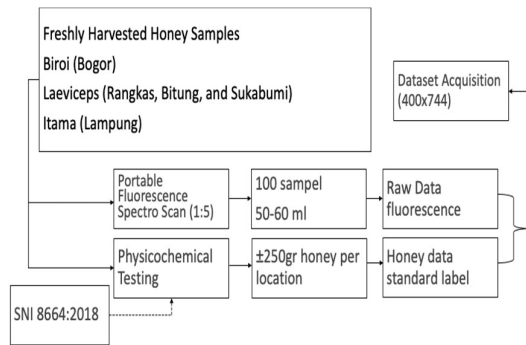


Figure 2. Stingless bee honey dataset acquisition

Figure 3 shows the research focused on identifying appropriate parameters for building an analytical prediction model from various primary data sources for the case of stingless honey in Indonesia. This research focused on developing spectral data processing and applying machine learning techniques to discover patterns/knowledge in the data and their visualization, which are useful for planning product purity quality detection. Furthermore, this research also focused on implementing the model in the form of mobile application-based software that can be used by users, taking into account aspects of user friendliness and user experience (UI/UX).

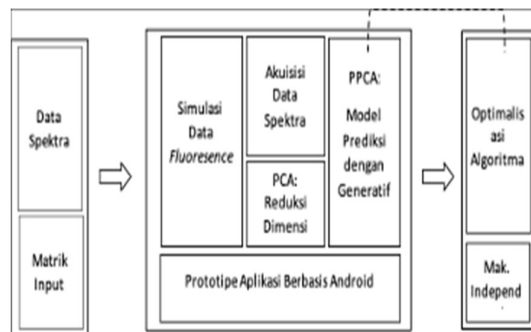


Figure 3. Algorithm Implementation Stages

The model is created from the best clustering results from the pre-processing stage, then classification algorithms such as Support Vector Machines, Random Forest, K-Nearest Neighbor, and Naive Bayes are applied. The resulting model is implemented in the development of an Android-based application using the Scrum Agile methodology.

The sprint concludes with a consultation held with stakeholders involving specific experts, product owners, and the production team [13]. So that it results in the development of a model with feature correlation and non-redundant, by sorting out the most important

features that are run [14]. User evaluation was conducted on experts and was carried out using direct interview methods to confirm the classification results produced by the model. All classes are represented in each test scenario. Primarily, all measurements (i.e., spectra) of samples are assigned to either the training data setting (70%) or the test data setting (30%) to ensure that the test setting does not include data from the same samples as the model was trained [15].

The overall dataset is 400x744, but after special preprocessing, it becomes 400x263. This dataset size is then used in the data split method to differentiate between training and test data. However, these datasets were taken from the same sample data collection.

### 3. RESULTS AND DISCUSSION

This section presents the final implementation results of the geographic indication classification model development, resulting in a prototype mobile application. Previously, the initial backend development of a cloud-based application with a 3-tier application architecture was conducted with the goal of achieving good and stable performance [10]. This article then presents the next steps, including the development of information on the characteristics of stingless bee honey products, modeling the geographic indication classification to evaluate the best performance, and developing the mobile application.

#### 3.1. Information Characteristics

The correlation analysis between the results of clinical tests on honey content and the honey spectrum measurement data, in Table I, shows the correlation results based on the honey's harvesting origin. Product content detection uses fluorescence, with peak values indicating concentration in a single dominant parameter. This is demonstrated in Table I, where the peak spectrum value is at the highest correlation point.

Table 1. Honey characteristic information

No.	O	W	D	HMF	RS	S	A
1.	L	0.3	<b>0.6</b>	<b>0.9</b>	<b>0.6</b>	<b>0.6</b>	0.4
2.	B	0.3	<b>0.6</b>	0.3	0.4	0.3	0.3
3.	Sk	<b>0.6</b>	<b>0.7</b>	<b>0.7</b>	0.5	<b>0.6</b>	<b>0.7</b>

4. Rb 0.5 0.3 0.3 0.4 0.5 0.7

Botanical identification – Flower pollen as natural vegetation of bees, for the dominance of certain types of bee food [3]. In Table 1, the area of origin of honey or Origin is shown with the label 'O', there are four regions namely L = Lampung, B = Bogor, Sk = Sukabumi and Rb = Rangkasbitung. Lampung has the highest significance correlation value for the clinical features of Hydroxymethylfurfural (HMF) of  $0.863 \approx 0.9$ , Diastase (D) of  $0.609 \approx 0.6$ , Reducing Sugar (RS) of  $0.627 \approx 0.6$ , and Sucrose (S) of  $0.637 \approx 0.6$ . Bogor has the highest significance correlation value only for the Diastase (D) feature of  $0.577 = 0.6$ . Sukabumi (Sk) and Rangkasbitung (Rb), both are dominant in the Acidity (A) feature with correlation values of  $0.743 = 0.7$  and  $0.688 = 0.7$ , respectively. Sk is also dominant in several other features with correlation values above or equal to 0.6 in the water content (W), Diastase (D), Hydroxymethylfurfural (HMF), and Saccharose (S) features. While Rb above or equal to 0.5 is also dominant in the clinical features of water content (W) and Saccharose (S).

Table 2. Tukey's HSD Results

No.	Parameter	Average of Mean Diff	Category
1	Water Content	2.28	Medium
2	Diastase	1.78	Medium
3	Hydroxymethylfurfural	3.6	High
4	Reducing Sugar	0.01	Very Low
5	Saccharose	1.33	Medium
6	Acidity	1.85	Medium-High

Tukey's HSD was also conducted to determine which locations differed significantly from others. The results of the analysis of clinical trial parameters data using the Tukey's HSD method in Table 2 show that the average difference between honey content is significantly influenced by honey origin. The results of this relationship are indicated by the difference between each honey origin and its dominant content. The highest significance value with the highest average difference of 3.6 for the Hydroxymethylfurfural (HMF) parameter. The significance value with the medium category has an average difference of 2.28, 1.78, and 1.33 for the Water Content, Diastase, and Saccharose parameters,

respectively. Meanwhile, the significance value with the medium-high category has an average difference of 1.85 for the Acidity parameter.

### 3.2. Modeling

One of the output targets of this research is the application of a Geographical Indication classification model to a mobile programming-based application. This is one way to provide solutions that meet the target for downstream research. PCA has been found useful in various analytical techniques such as spectrometry, Raman, NIR, UV-Vis, fluorescence, and X-ray spectroscopy [16]. The advantage of this preprocessing method is that it analyzes spectral patterns using a parameter estimation approach that maximizes data clustering. It is also reliable in reducing noise, such as outliers and mismatched data groups, based on the resulting cluster structure [7][9]. Pearson correlation is used to measure the correlation between two variables and is defined as the quotient of the covariance and standard deviation between the two variables [17]. The first variable is the parameter value from the clinical test results, and the second is the value from the honey product spectrum measurement results.

#### 3.2.1. Model Evaluation

The results of the classification algorithm comparison are shown in Table 3, where the model produces excellent performance with an accuracy value above 0.95. This is the result of classification modeling using several classification algorithms such as Support Vector Machine (SVM) and K-Neighbors (KNN). Random Forest (RF) and Naïve Bayes (NB) can achieve an accuracy of 0.99.

Model accuracy was achieved through several comprehensive pre-processes carried out in a planned manner through ongoing research scenarios. Based on the model evaluation, K-Fold Cross Validation (CV) was also performed, which showed a difference between the average CV and training accuracy of approximately 1%, indicating no overfitting and excellent generalization by the model. Furthermore, underfitting also did not occur, as the accuracy of the training data and CV accuracy were at the same high level.

**Table 3.** Evaluation of GI classification model

No.	Algorithm	Precision	Accuracy	Recall
1.	SVM ( <i>support vector machine</i> )	0.9545	0.9500	0.9444
2.	RF (random forest)	0.9917	0.9875	0.9762
3.	KNN (K-Neighbors Classifier)	0.9583	0.9750	0.9524
4.	NB (Gaussian Naïve Bayes)	0.9917	0.9875	0.9762

Evaluation of each algorithm showed a recall value above 0.94, indicating that the model correctly classifies the data, according to the origin or class label in the dataset. This was achieved after pre-processing, feature selection, which successfully reduced the remaining data columns to only 273, and dimensionality reduction, which reduced the data dimension to 3 Principal Components (PC) while maintaining a cumulative variance value equal to or greater than 95%.

This study uses a generative modeling approach, in this case the Probability Principal Component Analysis (PPCA) algorithm. PPCA is explained in this article in general terms, determining the number of PCs based on the dataset's representativeness and maintaining a cumulative variance of 95%. This differs significantly from the standard PCA approach in discriminant models. More details are described in a separate article, also part of this study.

### 3.2.2. Performance Analysis

The model performance measurement method is carried out using the Benchmarking method approach with the aim of comparing the best machine learning models [18]. This method is carried out by several initial settings such as the similarity of CPU and GPU settings, the similarity of parameter configurations and the precision and recall values of the model. In Table 4, it is shown that the model with the

fastest performance in terms of runtime reporting results in micro-seconds is the model with the SVM and NB algorithms, namely 1,685 ms and 1,702 ms. However, the results of the precision and recall values of the model obtained in Table 3, show that RF and NB are the best with values of 0.9917 and 0.9762. This indicates the best model with a dataset that has undergone pre-processing according to the proposed research, resulting in the NB algorithm having the best classification performance. More effective results in utilizing the most important and represented dataset features with smaller dimensions significantly improve application performance.

**Table 4.** Classification model performance

No	Algorithm	Configure	Time (ms)
1.	SVM	SVC()	1.685
2.	RF	RandomForestClassifier()	14.977
3.	KNN	KNeighborsClassifier()	6.122
4.	NB	GaussianNB()	1.702

Table 4, shows that the classification algorithm used has a performance in micro-seconds, and the fastest is SVM at 1.685ms. The model was created without adding any parameters in order to determine the effectiveness of the performance of the pre-processing algorithm used in this study.

But overall, based on the performance measurement method used, it was concluded that NB was the best with a runtime value of 1,702 ms and precision and recall values above 0.99.

### 3.3. Application

Each product developed in each sprint can be identified and submitted as a requirement for the next sprint [19].

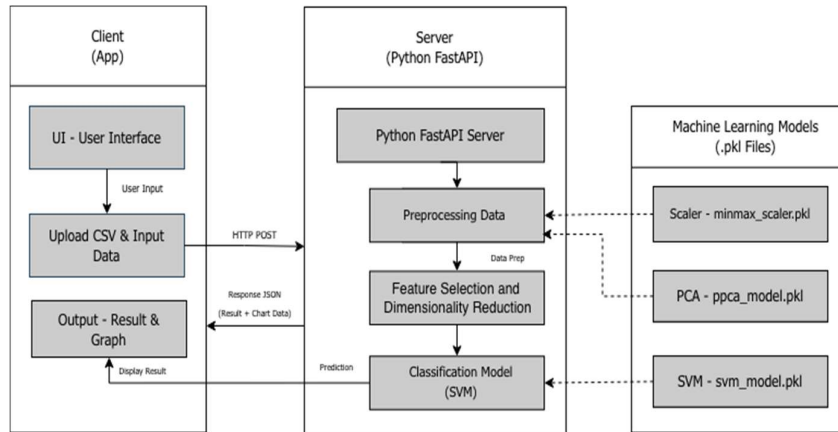


Figure 4. Client-Server Apps Architecture

The mobile-based application development, conducted using the Scrum methodology in stages, resulted in the release of an initial mock-up that only conveyed the results of the honey product region classification. In Figure 4, the Client-Server Apps Architecture is an implementation of mobile-based application development for geographic indication-based classification models. The three main parts are the Server Side with Python FastAPI, the Client Side for Apps, and the application backend on the AI model. The Server with FastAPI plays a role in integrating Machine Learning (AI Models) at every main stage of modeling, such as data preprocessing, feature selection and dimensionality reduction, and the use of classification algorithms. Integration is carried out by providing model prediction results and several graphical displays for display in Apps with JSON responses. HTTP POST is used to ensure the trigger input of the dataset file to be

processed is in accordance with the data acquisition provisions.

The client's considerations prioritize feasibility testing and suitability to field needs. In this case, the development documentation used was User Acceptance Test (UAT). Testing was conducted on every farmer at the honey sampling location from stingless bees. The results formed the product backlog used by the development team to develop a mock-up or prototype of the application.

The classification results from each algorithm accurately classified honey types with very few misclassifications. This is a result of the pre-processing stage, which successfully addressed noise in determining the label/class of honey products. Smartphones, with their many digital imaging features, are now becoming an excellent analytical platform for the development of point-of-care testing sensors [1][20].

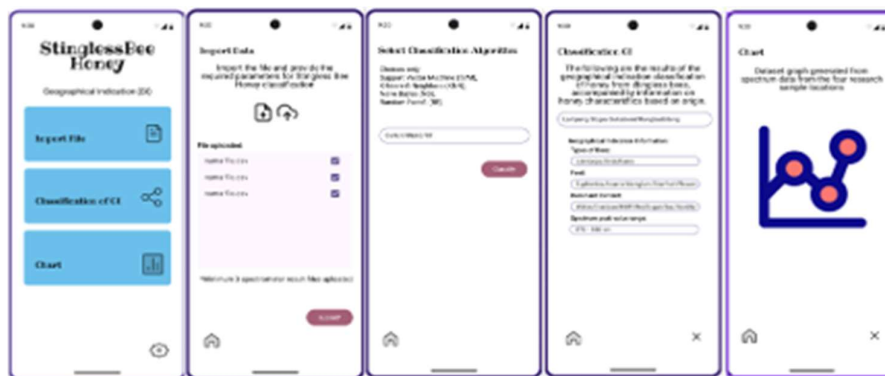


Figure 5. Application mock up

Figure 5, a mock-up of the final application, shows the first page showing the file import feature for honey spectrum data input, the geographic indication classification feature, and the feature for displaying spectrum graphs from the dataset used.

HoneyLyze is a honey origin prediction application that provides information on dominant constituent characteristics and key spectral values. It also provides information on the type of bee (entomologic) and dominant forage (vegetation/botanical) based on the predicted origin.

The application was developed based on observations of the conditions of the farmers and the specifications of the portable spectrophotometer used. Interviews with the farmers revealed diverse understandings of stingless beekeeping, but a common thread is the importance of consumer perceptions of the authenticity of the harvested honey. Therefore, this application demonstrates the initial steps in predicting honey authenticity, which involves classifying the geographical indications of honey products from stingless bees.



Figure 6. Application start page

The homepage of the honeylyze app takes users to upload new measurement data

from a portable spectrophotometer. The page displays the app's logo, specifically for honey product prediction, in this case, stingless bee honey. Figure 6 demonstrates the technology used to upload new datasets by simply clicking and searching for CSV data stored on a smartphone.

The next step button directs users to the next step after uploading a CSV dataset. The application is built with a data input trigger that then activates integration with the server to provide a JSON response to the classification results. This Geographical Indication Classification process uses spectrum measurements and clinical test results to determine the predicted honey's origin.

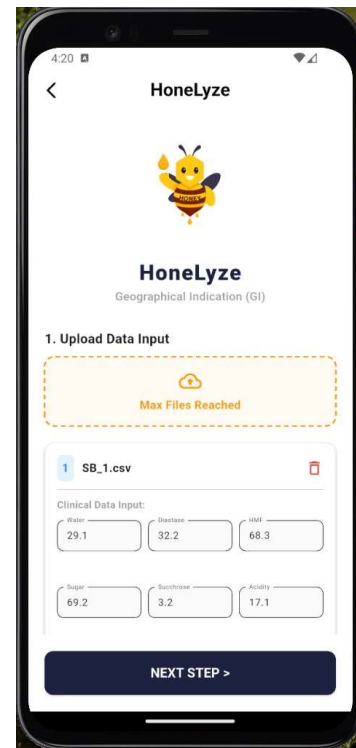


Figure 7. Clinical trial results data input page

Figure 7 shows the CSV file upload process being performed multiple times, depending on the classification model's requirements. Each time a file is uploaded, the app automatically prompts users to enter clinical test results. These can be entered using the clinical test results based on the region from which the laboratory results were obtained in this study. This is done if the user has not yet performed laboratory testing on their honey product.

This process is carried out following the results of the pre-processing stage of the classification model. The model requires input from a new dataset that is not limited to a single feature, this is done because the feature dimensionality reduction process is carried out before the data is tested or its classification is predicted. In Figure 7, input is carried out for six selected parameters according to the Indonesian National Standard for honey products. This application focuses on developing an effective and efficient backend because the spectrum data type has feature dimensions that can expand or become larger depending on the wavelength used. Technology development in the FastAPI-based HoneyLyz application for the server side facilitates the integration of machine learning model implementation.

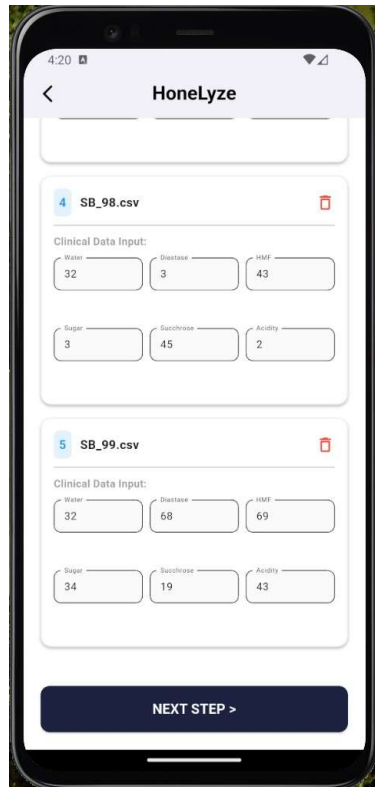


Figure 8. Spectrum measurement results data input page

The spectrum measurements used in Figure 7 and Figure 8 are the results of data acquisition starting from limiting the spectrum wavelength, selecting features with the highest correlation value (according to the specified threshold value), and reducing the dimension of features represented by three principal

components (PC). Overall, the series is carried out in a pre-process with a silhouette reference value parameter above 0.5 for the results of cluster structures or data groups that are free from overlap and have the best distance between groups.

Each new data upload automatically displays tabular input for clinical trial data. The app allows up to five new datasets. Users simply input clinical trial results into the tabular data columns provided, matching the laboratory results and test standards. The classification model used in the app incorporates a comprehensive data normalization process to facilitate model accuracy.

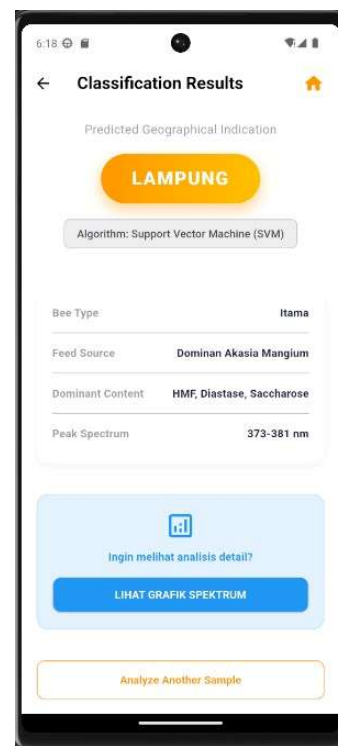


Figure 9. Honey classification results page

The honey classification results in Figure 9 show complete geographic information, beginning with the predicted origin of the honey. Lampung, as an example, uses newly inputted data to predict the bee species as Biroi and the primary vegetation as Euphorbia. Furthermore, the dominant content is water, with a peak spectral value at 375-721 nm. All of this information represents the detected characteristics of honey classified as geographically indicated.

Table 5. Validation of expert interview results

No	Jenis Lebah <i>stingless</i>	Lokasi	Nektar (Vegetasi)	Parameter dan Nilai Spektrum Dominan
1	Tetragonula Biroi	Bogor (CV. Nutrima)	Euphorbia	HMF, Diastase dan Reducing Sugar (427-455 nm) Blue Region
2	Tetragonula Laeviceps	Rangkas Bitung (CV. Nabila Natural)	Batavia, dan Bunga belimbing	Acidity, Saccharose dan HMF (427-458 nm) Blue Region
3	Heterotrigona Itama	Lampung (PT. Suhita-Farm 8 Way Kanan)	Dominan Akasia Mangium	HMF, Diastase dan Saccharose (373-381 nm) UV Region
4	Tetragonula Laeviceps	Sukabumi (futrabungsu Trigona)	Dominan Kalendra	Acidity, HMF dan Diastase (497-516 nm) Green Region

Table 5 shows the results of expert validation of the implemented model classification output. The application output image shows the classification of the origin of the harvest location "Bogor" showing the characteristics of the information owned that are in accordance with the expert validation in Table 5. The information in question is the type of bee "Biroi", the dominant clinical parameter "Water" and the peak spectrum value in the range of 375-721 nm. Overall, the output has been validated with information from experts, so the application has added value in the feasibility study for users.

Spectrum graphs are frequently used by spectrophotometer users, both benchtop and portable, to provide an indication of the analysis results of the product's content characteristics. One way to do this is by using the peak values of the graph at a specific wavelength range to indicate whether the measurement results are accurate or whether more comprehensive adjustments are needed. This application offers a similar alternative: by simply using the detailed analysis feature, users can compare the graph results for each sample origin.

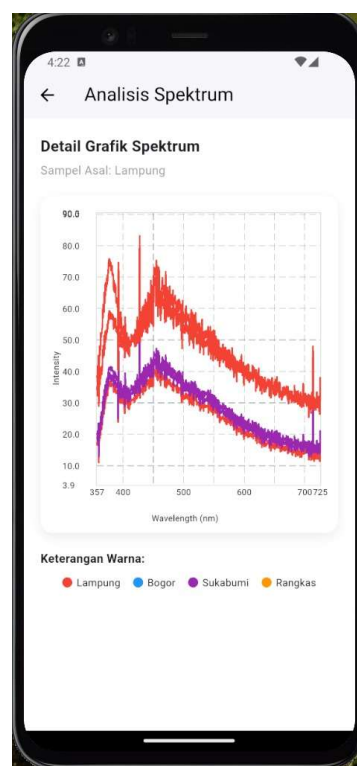


Figure 10. Honey spectrum graph results page

The development of this application is complemented by presenting the results of the spectrum graph from the results of the spectrum measurements on each honey sample used.

Figure 10 shows the spectrum graphs of honey samples from Lampung and Sukabumi. The resulting graphs show red for Lampung honey and purple for Sukabumi honey. The two graphs overlap and have similar peaks. This is a condition that is resolved by the proposed preprocessing algorithm and implemented in the application in this study.

## CONCLUSION

The characteristics of honey products from stingless bees are demonstrated as a Geographical Indication (GI) application by implementing a machine learning model. The model's accuracy is above 95%. This value is supported by the results of overfitting and underfitting evaluations, which yielded zero results. The Geographical Indication (GI) application for stingless bee honey is a honey product classification application based on honey spectrum measurements and clinical content testing based on six key laboratory parameters. The best GI classification model performance using the benchmarking method was 1,702 microseconds using the Naive Bayes (NB) classification algorithm.

This study successfully implemented the best GI classification model in a cloud-based application on the Android platform. Based on the results of the application feasibility study, which showed a 90% success rate, users, in this case the SBH cultivation industry, were concluded to be ready to implement and gradually eliminate doubts to reassure the public with the GI information provided in the application.

## ACKNOWLEDGEMENTS

We would like to thank the Ministry of Higher Education, Science and Technology for assistance in this research, under the Research Grant program with contract number 027/E5/PG.02.00.PL/2024 for the Doctoral Dissertation Research (PDD) scheme.

## REFERENCES

- [1] G. Chen, Y. Huang, and K. Chen, "Recent advances and applications of near infrared spectroscopy for honey quality assessment," *Adv. J. Food Sci. Technol.*, vol. 6, no. 4, pp. 461–467, 2014, doi: 10.19026/ajfst.6.55.
- [2] N. A. S. Suarin, K. S. Chia, and F. N. Kosmani, "Stingless Bee Honey Classification Using near Infrared Light Coupled with Artificial Neural Network," *2020 Zooming Innov. Consum. Technol. Conf. ZINC 2020*, pp. 99–102, 2020, doi: 10.1109/ZINC50678.2020.9161785.
- [3] A. Aziz, "FULL PAPER Adulterated stingless bee honey identification using VIS - NIR spectroscopy," vol. 5, pp. 85–93, 2021.
- [4] H. A. Hidalgo, A. R. Nicolas, and R. Cedon, "Development barriers of stingless bee honey industry in Bicol, Philippines," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, no. 3, pp. 1245–1251, 2020, doi: 10.18517/ijaseit.10.3.4747.
- [5] A. Hassoun, I. Mâge, W. F. Schmidt, H. T. Temiz, and L. Li, "Fraud in Animal Origin Food Products: Advances in Emerging Spectroscopic Detection Methods over the Past Five Years," *Foods*, vol. 9, pp. 1–41, 2020, doi: 10.3390/foods9081069.
- [6] M. Yulia and D. Suhandy, "Authentication of organic Lampung robusta ground roasted coffee by UV-visible spectroscopy and PLS-DA method," *J. Phys. Conf. Ser.*, vol. 1341, no. 2, 2019, doi: 10.1088/1742-6596/1341/2/022006.
- [7] D. Hou, S. Liu, J. Zhang, F. Chen, P. Huang, and G. Zhang, "Online monitoring of water-quality anomaly in water distribution systems based on probabilistic principal component analysis by UV-vis absorption spectroscopy," *J. Spectrosc.*, vol. 2014, 2014, doi: 10.1155/2014/150636.
- [8] I. E. Bell and G. V. G. Baranoski, "Reducing the dimensionality of plant spectral databases," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 570–576, 2004, doi: 10.1109/TGRS.2003.821697.
- [9] Y. Zhu, S. Zhao, Y. Zhang, C. Zhang, and J. Wu, "A Review of Statistical-Based Fault Detection and Diagnosis with Probabilistic Models," *Symmetry (Basel)*, vol. 16, no. 4, pp. 1–23, 2024, doi: 10.3390/sym16040455.
- [10] Y. A. Purwanto, H. Maulana, S. Hartono W, H. Sukoco, and Diding Suhandy, "Cloud-Based Predictive Mobile Application for Assessing Honey Purity From Stingless Bees," *J. Ilm. Kursor*, vol. 12, no. 4, pp. 191–200, Dec. 2024, doi: 10.21107/kursor.v12i4.420.
- [11] M. J. Aliaño-González, M. Ferreira-González, E. Espada-Bellido, M. Palma, and G. F. Barbero, "A screening method

- based on Visible-NIR spectroscopy for the identification and quantification of different adulterants in high-quality honey,” *Talanta*, vol. 203, pp. 235–241, 2019, doi: 10.1016/j.talanta.2019.05.067.
- [12] A. Priyadi and M. D. E. Wiratmoko, “Karakteristik Fisikokimia Madu Heterotrigona itama Asal Provinsi Riau,” *Wahana For. J. Kehutan.*, vol. 18, no. 2, pp. 13–28, 2023, doi: 10.31849/forestra.v18i2.11107.
- [13] L. Neelu and D. Kavitha, “Estimation of software quality parameters for hybrid agile process model,” *SN Appl. Sci.*, vol. 3, no. 3, pp. 1–11, 2021, doi: 10.1007/s42452-021-04305-0.
- [14] H. Chen, Z. Lou, and G. Shen, “An integrated flexible multifunctional sensing system for simultaneous monitoring of environment signals,” *Sci. China Mater.*, vol. 63, no. 12, pp. 2560–2569, 2020, doi: 10.1007/s40843-020-1376-4.
- [15] H. Parastar, G. van Kollenburg, Y. Weesepeel, A. van den Doel, L. Buydens, and J. Jansen, “Integration of handheld NIR and machine learning to ‘Measure & Monitor’ chicken meat authenticity,” *Food Control*, vol. 112, no. January, p. 107149, Jun. 2020, doi: 10.1016/j.foodcont.2020.107149.
- [16] J. R. Beattie and F. W. L. Esmonde-White, “Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra,” *Appl. Spectrosc.*, vol. 75, no. 4, pp. 361–375, 2021, doi: 10.1177/0003702820987847.
- [17] Z. Wang, T. Wang, Y. Yang, X. Mi, and J. Wang, “Differential Confocal Optical Probes with Optimized Detection Efficiency and Pearson Correlation Coefficient Strategy Based on the Peak-Clustering Algorithm,” *Micromachines*, vol. 14, no. 6, 2023, doi: 10.3390/mi14061163.
- [18] V. E. Baciú, J. Stiens, and B. da Silva, “MLino bench: A comprehensive benchmarking tool for evaluating ML models on edge devices,” *J. Syst. Archit.*, vol. 155, no. June, p. 103262, 2024, doi: 10.1016/j.sysarc.2024.103262.
- [19] N. Naik, P. Jenkins, and D. Newell, *Learning Agile Scrum Methodology Using the Groupware Tool Trello*  $\text{\textregistered}$  *Through Collaborative Working*, vol. 993. Springer International Publishing, 2020, doi: 10.1007/978-3-030-22354-0\_31.
- [20] Y. Shen, Y. Wei, C. Zhu, J. Cao, and D. M. Han, “Ratiometric fluorescent signals-driven smartphone-based portable sensors for onsite visual detection of food contaminants,” *Coord. Chem. Rev.*, 2022, doi: doi.org/10.1016/j.ccr.2022.214442.