OPEN ACCESS
Vol. 7, No. 2, pp.183-192, May 2025

ISSN : 2622-8106 (ONLINE)
DOI: 10.20895/inista.v7i2.1832

Journal of Informatics, Information System, Software Engineering and Applications
(INISTA)

Class Balancing and Parameter Tuning of Machine Learning Models for Enhancing Aphrodisiac Herbal Plant Classification

Puguh Jayadi *[1], Weka Sidha Bhagawan [2], Jofanza Denis Aldida [3]

[1, 3]*Informatics Study Program, Faculty of Engineering, PGRI Madiun*
[2]*Pharmacy Study Program, Faculty of Health Sciences and Sciences, Universitas PGRI Madiun*
[1,2,3] *Jl. Auri No.14-16, Kanigoro, Kartoharjo, Madiun City, 63117, East Java, Indonesia*

[1] puguh.jayadi@unipma.ac.id
[2] weka.sidha@unipma.ac.id
[3] jofanza_2205101036@mhs.unipma.ac.id

Received on 04-05-2025, revised on 18-05-2025, accepted on 24-05-2025

Abstract

Herbal plants with aphrodisiac claims are an important part of traditional medicine that continues to evolve within the modern scientific context. However, the classification process for these plant claims is often done manually and subjectively, necessitating a more objective, data-driven approach. Artificial Intelligence (AI) and its various derivatives, such as Machine Learning, present a reliable solution for several related classification studies. The primary challenge in classification lies in data class imbalance and selecting the optimal model parameters. This study proposes an integrated approach that utilizes machine learning algorithms, including Random Forest, Support Vector Machine (SVM), and XGBoost, combined with SMOTE class balancing techniques and hyperparameter tuning through Grid Search, Random Search, and Bayesian Optimization. Experiments were conducted on a dataset of herbal plants with attributes and labels of aphrodisiac claims, and the results were evaluated based on accuracy, precision, recall, and execution time. The findings indicated that the combinatorial approach significantly improved model performance compared to the basic approach. Among the hyperparameter tuning results, the SVM method achieved the best accuracy (0.889) and precision (0.889). This research contributes to the development of an AI-based classification system in the field of ethnopharmacology. It can serve as a reference for creating scientifically validated databases of herbal plants.

Keywords: Classification of Herbal Plants, Aphrodisiacs, Machine Learning, Hyperparameter Tuning.

Corresponding Author:
*Puguh Jayadi
Informatics Study Program, Faculty of Engineering, PGRI Madiun
Jl. Auri No.14-16, Kanigoro, Kartoharjo, Madiun City, 63117, East Java, Indonesia
Email: puguh.jayadi@unipma.ac.id

I. INTRODUCTION

Herbal plants have become an integral part of traditional medicine practices in many cultures worldwide, including Indonesia, which is renowned for its extraordinary biodiversity. Thousands of plant species have been used for generations by the community to treat various diseases, strengthen vitality, and improve quality of life [1], [2], [3]. Among the most prominent benefit claims is the aphrodisiac effect, which can increase arousal and sexual function [4], [5]. Plants believed to have this effect are often incorporated into traditional herbal remedies, which are consumed in the form of extracts formulated into modern health products [6], [7], [8]. However, the labeling and classification of these claims are still performed manually and subjectively, thus posing challenges in terms of scientific validity and standardization of classification [9].

In the era of digitalization and advances in Artificial Intelligence (AI) technology, data-driven approaches are starting to be widely used to support scientific validation of traditional information, including claims of the benefits of herbal plants [10], [11]. One of the most promising approaches in this regard is the application of Machine Learning techniques, which allow the modeling of complex relationships between features without explicitly defining their mathematical functions [12]. Algorithms such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Decision Tree, and XGBoost have been proven to be effective in classification tasks, including in the fields of biomedicine, digital agriculture, and plant pharmacology [13], [14]. The advantage of this approach lies in its ability to learn from data and make predictions based on the statistical patterns stored within it [15], [16].

However, one of the main problems in the application of Machine Learning for the classification of herbal plants, especially the claim of aphrodisiac effects, is the data imbalance (class imbalance) [17], [18]. In most cases, the available data tends to dominate one or two specific classes, while the rest have very low representation. This imbalance causes the model to become biased towards the majority class, and its performance in recognizing plants with minority claims becomes abysmally poor [19], [20]. In addition, model performance is also affected by the configuration of parameters (*hyperparameters*), which, if not set correctly, can significantly reduce the accuracy, precision, and generalization ability of the model [21], [22]. Therefore, these two problems must be addressed methodologically to ensure the classification results are more representative and reliable.

Several previous studies have demonstrated that resampling approaches, such as the Synthetic Minority Oversampling Technique (SMOTE), effectively address the problem of class imbalance by generating synthetic samples that closely approximate the distribution of minority classes. For example, research by Chawla et al. (2002) introduced SMOTE as a solution that significantly improves classification performance compared to traditional undersampling approaches [22]. On the other hand, the hyperparameter tuning process has also been empirically proven to have a direct impact on the final performance of the model. For example, a study by Bergstra and Bengio (2012) showed that Random Search can explore parameter spaces more effectively than Grid Search in many scenarios [23]. Then, Bayesian Optimization offers a more efficient approach by modeling objective functions using probability to focus the search on the most promising parameter regions. However, few studies have integratively applied class balancing and hyperparameter optimization strategies simultaneously to the classification of herbal plants, especially in specific contexts such as those related to aphrodisiac claims.

The knowledge gap that emerged from these studies was the lack of an integrated approach that simultaneously addressed two critical aspects: class imbalance and model parameter tuning. In the context of classifying herbal plants, especially as it relates to aphrodisiac claims, precision is of great importance because misclassification can have implications for the use of inappropriate or therapeutically ineffective plants. Therefore, research that focuses on selecting the correct algorithm and systematically integrates class balancing strategies and parameter optimization is needed to build a more accurate and reliable classification system.

This research will address this need by developing a machine learning-based approach for classifying aphrodisiac herbal plant claims, which combines SMOTE class balancing techniques with hyperparameter tuning using Grid Search, Random Search, and Bayesian Optimization. By comparing the performance of several commonly used classification algorithms, this study aims to identify the optimal combination of techniques and parameters that improves accuracy, precision, recall, and processing time efficiency. The focus on this evaluation indicator is very relevant, considering the importance of scientific validation of herbal plant claims in supporting the development of safe and effective herbal medicines.

With this approach, this research aims to contribute to the development of an artificial intelligence-based classification system in plant ethnopharmacology and bioinformatics. In addition, the results of this study can serve as a starting point for the development of a more comprehensive, data-driven automatic classification system, which relies not only on ethnobotanical narratives but also incorporates predictive analysis that can be scientifically accounted for.

II. RESEARCH METHOD

This research was conducted through the following main stages: data acquisition, preprocessing, class balancing, development and tuning of machine learning models, and model performance evaluation, as illustrated in Figure 1. The dataset comprises a collection of aphrodisiac herbal plants from various provinces in Indonesia, including taxonomic attributes (such as Family, Species, and Local Names) and geographic information (Province, County), as shown in Table 1. The dataset was collected from several research studies related to herbal plants [24], [25], [26]. The data consists of 466 rows and six features. Target labels (Claims) are classified into several categories of medical claims that are discrete and serve as dependent variables in the classification task [27], [28].

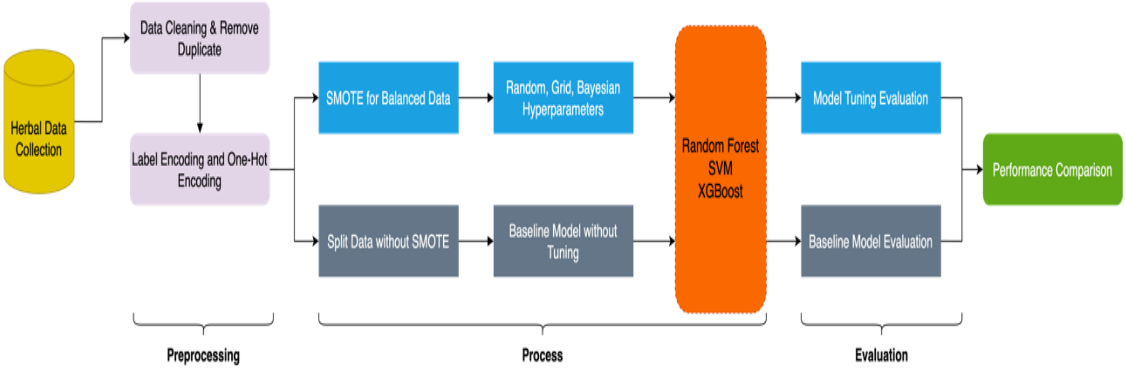

Fig. 1. Research Methodology

TABLE I. DATASET DESCRIPTION

Attribute	Datatype	Description
Family	String	Taxonomic families of herbal plants, such as *Acanthaceae*, *Zingiberaceae*, etc.
Species	String	Scientific species names of herbal plants are written in binomial format, such as *Acorus calamus L.*
Local Name	String	Local or common names of herbs in the local area.
Province	String	Provinces where the plant is used as an herb.
Regency	String	Regency/city where the plant is used as an herbal medicine.
Claims	String	Claims of the benefits of herbal plants associated with aphrodisiacs or male reproductive organs.

The preprocessing stage begins with the conversion of categorical attributes to numerical forms. The data cleanup process involves removing blank values and duplicate data. A label encoding technique converts each unique class into an integer for the Claims target variable. Mathematically, this process can be written as an injective function in Equation (1) [15]. With C is a set of aphrodisiac claims, class k is the total sum of unique claims, and f is the mapping function that associates each claim c_i integers are not negative.

$$f : C \to \mathbb{Z}, \quad f(c_i) = i - 1, i = 1, 2, \ldots, k \tag{1}$$

Furthermore, categorical predictor attributes, such as family, species, local name, province, and regency, are converted using One-Hot Encoding (OHE). This representation forms a dimensional binary vector $|A|$, where each category value is represented by a vector with one value of 1 and the rest of the values being 0. This representation can be written on Equation (2), with x_j The category value j of the A, and the result vector has a value of 1 in position j [15].

$$\text{OHE}(x_j) = [0,0, \ldots, 1, \ldots, 0] \tag{2}$$

After the encoding process, the data is divided into two experimental paths. The first track uses original data (imbalanced), while the second uses the data from class balancing through the Synthetic Minority Over-sampling Technique (SMOTE). The SMOTE technique performs a linear interpolation between the minority data and its closest neighbors, generating synthetic data with Equations (3), with x_i a feature vector of minority data x_{zi} as a near neighbor of x_i, and λ is a random number from a uniform distribution $U\ (0, 1)$.

$$x_{\text{new}} = x_i + \lambda \cdot (x_{zi} - x_i), \lambda \sim U\ (0,1) \tag{3}$$

SMOTE was chosen because the dataset exhibits a disparity between frequent and infrequent claims. The oversampling process was carried out before the data was divided into training and test sets, with an 80:20 split ratio on each track. This approach aims to evaluate the impact of class balancing on the performance of the classification model.

In the next stage, three machine learning models were used to build a classification system: Random Forest, Support Vector Machine (SVM), and XGBoost. All three were chosen because they represent three different architectures. Random Forest represents a decision tree-based ensemble approach, and SVM relies on the maximal margin principle for class separation. At the same time, XGBoost is a robust boosting algorithm with regularization capabilities. Each model was implemented under two conditions: as a baseline without tuning using the original data and with hyperparameter tuning using data balanced via SMOTE. Random Forest employs the ensemble principle by combining the results of T decision trees [29]. The final prediction is determined by majority voting, expressed in Equation (4). Where $h_t(x)$ is a prediction from the tree t, and $\hat{y}$ is the result of the final prediction.

$$\hat{y} = mode(\{ h_t(x) \}_{t=1}^{T}) \tag{4}$$

SVM seeks an optimal separator hyperplane that maximizes the margin between the two classes. SVM optimization function as in Equation (5) [30]. With w is the weight vector b is biased, y_i is the target label of the observation i, and $\phi(x_i)$ is the transformation of a feature into a high-dimensional space through a Gaussian kernel.

$$\min_{w,b} \frac{1}{2} \parallel w \parallel^2 \; s.t \; y_i(w \,.\phi(x_i) + b \,) \geq 1 \tag{5}$$

For the XGBoost model, the learning process is based on the optimization of objective functions through regularization, aformulated in Equation (6) [31]. With l as a logarithmic loss function, $\Omega(f)$ as a regularization function, γ as a penalty of tree complexity, λ as a parameter of L2 regularization, and w as the weight of the leaves in the decision tree.

$$\mathcal{L}\,(\phi) = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \Omega\,(f_k)\,, \Omega(f) = \gamma\, T + \frac{1}{2}\lambda \parallel w \parallel^2 \tag{6}$$

To improve the classification accuracy, a hyperparameter search was conducted using three different approaches: Randomized Search CV, Grid Search CV, and Bayesian Optimization (BayesSearchCV). The Randomized Search approach explores the parameter space randomly in a limited number of iterations, while Grid Search performs a thorough search of the entire specified parameter combination. At the same time, Bayesian Optimization uses a probabilistic approach to redirect searches to more promising parameter areas. The tuned parameters are adapted to the characteristics of each model. For Random Forest, tuning is performed on the number of trees (n_estimators), maximum depth (max_depth), and the minimum number of samples required for separation (*min_samples_split*). The C, gamma, and kernel type parameters are optimized for SVM. For XGBoost, tuning involves adjusting the number of estimators, maximum depth, and learning rate. The entire tuning process employs a three-fold Stratified K-fold cross-validation strategy to maintain consistent class distributions in each fold, thereby making the model evaluation more robust. For Bayesian Optimization, parameters are selected based on acquisition functions such as Expected Improvement, which is formulated as in Equation (7) with $f(x)$ as a surrogate model (e.g., Gaussian Process) and $f(x^{+})$ is the highest value that has been found so far.

$$x_{t+1} = arg \max_{x} \mathbb{E}\; [\max(0, f(x) - f(x^{+}))] \tag{7}$$

The performance of each model was evaluated using the main classification metrics, precision, recall, and F1-score. All metrics are calculated based on a weighted average for class imbalances. Evaluation is done on test data not involved in the training process or oversampling [32], [33]. The review results are obtained through the classification_report function of Scikit-learn and stored in a structured format for comparative analysis. Evaluation is carried out using precision, recall, and F1-score metrics in Equations (8), (9), (10), whereTP_i, FP_i dan FN_i Each is the number of *True Positives*, *False Positives*, and *False Negatives* in the class i, and K is the total number of classes [17], [18].

$$\text{Precision} = \frac{\sum_{i=1}^{k} \mathrm{TP}_i}{\sum_{i=1}^{k} TP_i + FP_i} \tag{8}$$

$$\text{Recall} = \frac{\sum_{i=1}^{k} \mathrm{TP}_i}{\sum_{i=1}^{k} \mathrm{TP}_i + \mathrm{FN}_i} \tag{9}$$

$$\mathrm{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

$$\text{Execution Time} = t_{finish} - t_{start} \tag{11}$$

In addition, to evaluate the efficiency of the tuning method used, the computational time of each approach is recorded using the *time()* function. This allows for a trade-off analysis between improved accuracy performance and computational time load, particularly relevant in real-world model applications. The execution time of each tuning method is also recorded as a metric of computational efficiency, as shown in Equation (11). Where t_{finish} and t_{start} Each is the final timestamp, and the tuning process begins in seconds.

III. RESULTS AND DISCUSSION

Table II presents the performance evaluation results of three Machine Learning algorithms — Random Forest, Support Vector Machine (SVM), and XGBoost — on the classification of aphrodisiac herbal plant claims. Each algorithm was evaluated based on three main metrics: *precision*, *recall*, and *F1-score*, as well as execution time (in seconds), both at baseline and after applying three *hyperparameter tuning techniques*: RandomizedSearchCV, GridSearchCV, and BayesSearchCV.

TABLE II. RESULT MATRIX COMPARISON

Hyperparameter	Matrix	Random Forest	SVM	XGBoost
Baseline	Precision	0,548	0,376	**0,562**
	Recall	0,559	**0,613**	0,591
	F1-score	0,49	0,466	**0,552**
	Execution Time (s)	**0,359**	1,491	2,932
RandomizedSearchCV	Precision	0,832	**0,889**	0,788
	Recall	0,828	**0,889**	0,783
	F1-score	0,828	**0,888**	0,784
	Execution Time (s)	17,092	**5,257**	16,568
GridSearchCV	Precision	0,832	**0,889**	0,788
	Recall	0,828	**0,889**	0,783
	F1-score	0,828	**0,888**	0,784
	Execution Time (s)	**6,135**	8,154	19,579
BayesSearchCV	Precision	0,824	**0,836**	0,773
	Recall	0,811	**0,839**	0,772
	F1-score	0,81	**0,836**	0,772
	Execution Time (s)	**7,998**	16,496	15,589

The visualization in Figure 2 shows that the three tuning methods provide a significant improvement over the baseline. For example, the precision value in Random Forest increased from 0.548 to 0.832 after implementing RandomizedSearchCV. Similar improvements are also observed in the recall and F1-score, as shown in Figures 3 and 4. For SVM, GridSearchCV and RandomizedSearchCV managed to maintain recall values above 0.88, reflecting the model's ability to identify almost all relevant aphrodisiac claims. XGBoost also increased F1-score after tuning, although not as high as the other two algorithms.

In terms of execution time, as shown in Figure 5, the baseline consistently takes the shortest time because it does not involve a parameter search process. However, the RandomizedSearchCV and BayesSearchCV techniques provide a pretty good trade-off between accuracy and time efficiency. For example, GridSearchCV on XGBoost takes almost 20 seconds, whereas BayesSearchCV delivers competitive results in about 15 seconds.

The results obtained from this study show that the application of the data balancing technique (*class balancing*) using the Synthetic Minority Oversampling Technique (SMOTE) substantially improves the classification performance of aphrodisiac herbal plant claims. Before SMOTE, the dataset's class distribution was highly unbalanced, which caused the model to tend to be biased towards the majority class. This imbalance is evident from the relatively low values of precision and recall in the baseline model, particularly in the Support Vector Machine (SVM) and Random Forest models.

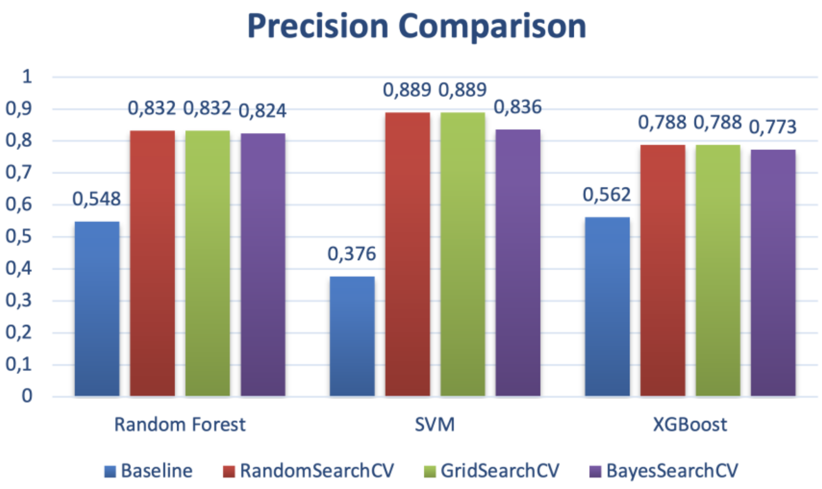

Fig. 2. Precision Comparison

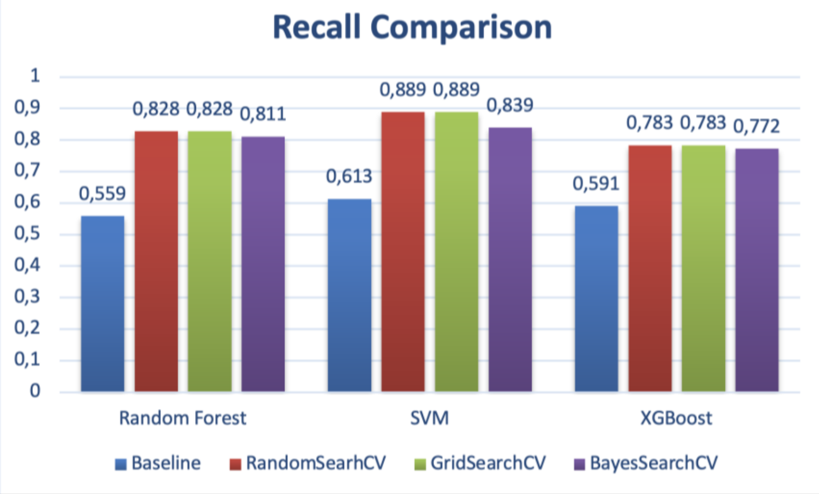

Fig. 3. Recall Comparison

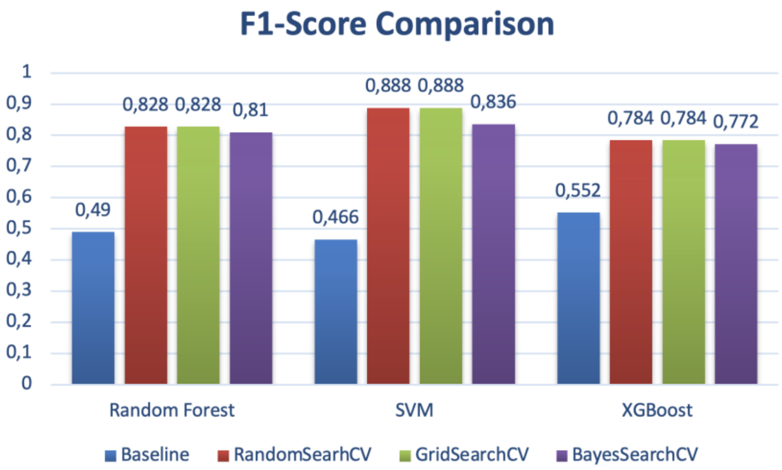

Fig. 4. F1-Score and Execution Time Comparison

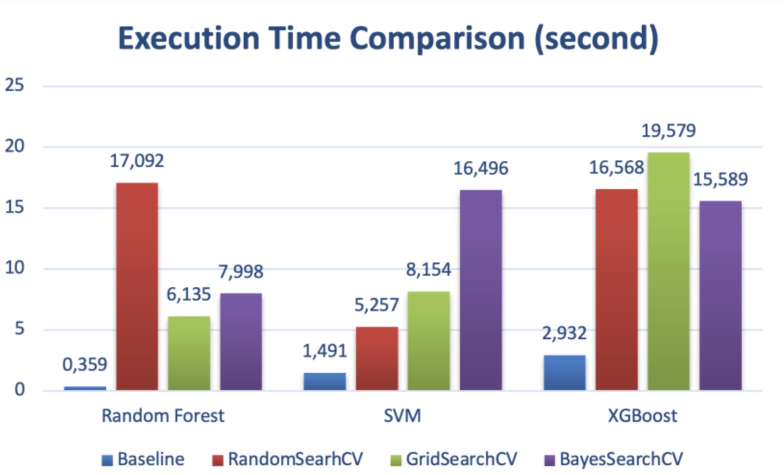

Fig. 5. Execution Time Comparison

After the data balancing process, the hyperparameter tuning stage was carried out using three different approaches: RandomizedSearchCV, GridSearchCV, and BayesSearchCV. Each approach contributes differently to improving model performance. RandomizedSearchCV has proven to be time-efficient and highly effective in finding optimal parameter combinations, as demonstrated by the increase in F1-scores across all three algorithms. This approach performs a random search of a predefined parameter space, thus reducing the computational burden compared to GridSearchCV, which tests all parameter combinations.

GridSearchCV, although it has higher execution times, provides more stable and precise results on SVMs. This follows the characteristics of the SVM algorithm, which relies heavily on selecting kernel parameters, as well as C and gamma values. The advantages of GridSearchCV in thoroughly exploring the parameter space give this algorithm an edge. However, the increased computing costs make this method less efficient for real-world application scenarios with limited resources.

BayesSearchCV, as a *Bayesian Optimization* method, demonstrates promising performance in terms of efficiency and effectiveness. Based on previous results, this technique leverages the acquisition function to explore the most promising parameter space adaptively. This approach yields competitive results, particularly on Random Forest, with a more efficient execution time compared to GridSearchCV. These findings align with the research of Hutter et al. (2019), which showed that Bayesian Optimization is effective in refining models with complex parameter spaces [21].

When evaluated based on precision and recall metrics, Random Forest and SVM achieve the most significant gains from parameter tuning. With its complex tree structure and sensitivity to the number of estimators and tree depth, Random Forest gains improved prediction accuracy once parameters such as n_estimators, max_depth, and min_samples_split are set appropriately. Meanwhile, in SVM, the C and gamma parameter adjustments greatly determine the optimal margin in the classification hyperplane.

The influence of data balancing and parameter tuning can also be evaluated from the perspective of execution time. Execution time increased significantly after implementing GridSearchCV, which aligns with the exponential computational complexity of the end-to-end parameter search. However, this improvement is still acceptable in the study context, especially when compared to the significant improvement in the classification evaluation metrics. This confirms the importance of the trade-off between execution time and predictive performance in applying machine learning algorithms for biological classification.

This study shows that the combination strategy of class balancing using SMOTE and hyperparameter optimization can significantly improve model performance. Bayesian-based parameter adjustment and other heuristic approaches have proven essential in enhancing the accuracy of classification systems, particularly in bioinformatics and ethnobotany, which often encounter data limitations and uneven distribution.

IV. CONCLUSION

This study successfully showed that the combination of *class balancing* techniques using SMOTE and *hyperparameter tuning* significantly improved the classification performance of aphrodisiac herbal plant claims. Machine learning models, such as Random Forest, Support Vector Machine, and XGBoost, have consistently improved in accuracy, precision, and recall after undergoing data balancing and parameter optimization. The application of the Bayesian Optimization method (BayesSearchCV) has proven to be efficient and effective in finding the optimal parameter configuration with lower execution time compared to GridSearchCV. These findings underscore the crucial role of precise preprocessing and parameter tuning in developing reliable classification models, particularly on frequently imbalanced biological datasets.

As a suggestion, future research could expand the scope of algorithms by integrating advanced ensemble approaches such as LightGBM or CatBoost for more optimal performance. Additionally, exploring other data balancing techniques, such as ADASYN or a combination of undersampling and oversampling, could be a focus of further development. Applying more complex cross-validation methods, such as stratified K-fold, is also recommended to ensure the model's generalization on new data. With a more comprehensive methodological approach, machine learning-based classification systems in the herbal domain can be developed into more accurate and reliable decision-making tools.

ACKNOWLEDGMENT

This research was funded by Lembaga Penelitian dan Pengabdian kepada Masyarakat (Institute for Research and Community Service) Universitas PGRI Madiun in 2025.

REFERENCES

[1] C.-T. Che, V. George, T. P. Ijinu, P. Pushpangadan, and K. Andrae-Marobela, “Traditional medicine,” in *Pharmacognosy*, Elsevier, 2024, pp. 11–28. doi: 10.1016/B978-0-443-18657-8.00037-2.

[2] N. Nasim, I. S. Sandeep, and S. Mohanty, “Plant-derived natural products for drug discovery: current approaches and prospects,” *The Nucleus*, vol. 65, no. 3, pp. 399–411, Dec. 2022, doi: 10.1007/s13237-022-00405-3.

[3] N. Chaachouay and L. Zidane, “Plant-Derived Natural Products: A Source for Drug Discovery and Development,” *Drugs and Drug Candidates*, vol. 3, no. 1, pp. 184–207, 2024, doi: 10.3390/ddc3010011.

[4] G. Nouioura, B. Lyoussi, and E. Derwich, “Rekindling desire: Unveiling the Aphrodisiac potential of Apiaceae Elixirs,” *Phytomedicine Plus*, vol. 4, no. 2, p. 100530, May 2024, doi: 10.1016/j.phyplu.2024.100530.

[5] A. M. Bubnova and A. V. Galchenko, “Natural Aphrodisiacs: Traditional Use, Mechanism of Action, Clinical Efficacy, and Safety,” *Nat Prod J*, vol. 14, no. 1, Feb. 2024, doi: 10.2174/2210315513666230324111231.

[6] I. Süntar, “Importance of ethnopharmacological studies in drug discovery: role of medicinal plants,” *Phytochemistry Reviews*, vol. 19, no. 5, pp. 1199–1209, Oct. 2020, doi: 10.1007/s11101-019-09629-9.

[7] M. Leonti, “The future is written: Impact of scripts on the cognition, selection, knowledge and transmission of medicinal plant use and its implications for ethnobotany and ethnopharmacology,” *J Ethnopharmacol*, vol. 134, no. 3, pp. 542–555, Apr. 2011, doi: 10.1016/j.jep.2011.01.017.

[8] F. Jamshidi-Kia, Z. Lorigooini, and H. Amini-Khoei, “Medicinal plants: Past history and future perspective,” *Journal of HerbMed Pharmacology*, vol. 7, no. 1, pp. 1–7, 2018, doi: 10.15171/jhp.2018.01.

[9] A. Andrade-Cetto and M. Heinrich, “From the field into the lab: Useful approaches to selecting species based on local knowledge,” *Front Pharmacol*, vol. APR, no. April, pp. 1–5, 2011, doi: 10.3389/fphar.2011.00020.

[10] S. Hanif *et al.*, “Integrative approach to the biochemical, and toxicological fingerprinting of Polygonum glabrum.: A computational and experimental synergy for a medicinal food plant,” *Food Biosci*, vol. 60, p. 104435, 2024, doi: https://doi.org/10.1016/j.fbio.2024.104435.

[11] K. L. D. Viet, K. Le Ha, T. N. Quoc, and V. T. Hoang, “Medicinal Plants Identification Using Federated Deep Learning,” *Procedia Comput Sci*, vol. 234, pp. 247–254, 2024, doi: https://doi.org/10.1016/j.procs.2024.02.171.

[12] H. Bouakkaz, M. Bouakkaz, C. A. Kerrache, and S. Dhelim, “Enhanced classification of medicinal plants using deep learning and optimized CNN architectures,” *Heliyon*, vol. 11, no. 3, p. e42385, 2025, doi: https://doi.org/10.1016/j.heliyon.2025.e42385.

[13] J. Cao *et al.*, “Graphene enhances artemisinin production in the traditional medicinal plant Artemisia annua via dynamic physiological processes and miRNA regulation,” *Plant Commun*, vol. 5, no. 3, p. 100742, 2024, doi: https://doi.org/10.1016/j.xplc.2023.100742.

[14] M. Jiang *et al.*, “Genetic diversity of the Chinese medicinal plant Astragali Radix based on transcriptome-derived SSR markers,” *Electronic Journal of Biotechnology*, vol. 62, pp. 13–20, 2023, doi: https://doi.org/10.1016/j.ejbt.2022.12.001.

[15] Z.-Y. Zhao *et al.*, “Major specialized natural products from the endangered plant Heptacodium miconioides, potential medicinal uses and insights into its longstanding unresolved systematic classification,” *Phytochemistry*, vol. 228, p. 114259, 2024, doi: https://doi.org/10.1016/j.phytochem.2024.114259.

[16] D. T. N. Nhut, T. D. Tan, T. N. Quoc, and V. T. Hoang, “Medicinal plant recognition based on Vision Transformer and BEiT,” *Procedia Comput Sci*, vol. 234, pp. 188–195, 2024, doi: https://doi.org/10.1016/j.procs.2024.02.165.

[17] A. K. Shembo, S. S. Ayichew, I. Stiers, A. Geremew, and L. Carson, “Classification and ordination analysis of wild medicinal plants in Ada’a district, Ethiopia: Implication for sustainable conservation and utilization,” *Ecological Frontiers*, vol. 44, no. 4, pp. 809–819, 2024, doi: https://doi.org/10.1016/j.ecofro.2024.04.002.

[18] T. P. Tran, F. Ud Din, L. Brankovic, C. Sanin, S. M. Hester, and M. Duc Hoang Le, “Incremental and Zero-Shot Machine Learning for Vietnamese Medicinal Plant Image Classification,” *Procedia Comput Sci*, vol. 246, pp. 606–615, 2024, doi: https://doi.org/10.1016/j.procs.2024.09.469.

[19] B. R. Pushpa, N. S. Rani, M. Chandrajith, N. Manohar, and S. S. K. Nair, “On the importance of integrating convolution features for Indian medicinal plant species classification using hierarchical machine learning approach,” *Ecol Inform*, vol. 81, p. 102611, 2024, doi: https://doi.org/10.1016/j.ecoinf.2024.102611.

[20] T. Asafo-Agyei, Y. Appau, K. B. Barimah, and A. Asase, “Medicinal plants used for management of diabetes and hypertension in Ghana,” *Heliyon*, vol. 9, no. 12, p. e22977, 2023, doi: https://doi.org/10.1016/j.heliyon.2023.e22977.

[21] M. Feurer and F. Hutter, “Hyperparameter Optimization,” 2019, pp. 3–33. doi: 10.1007/978-3-030-05318-5_1.

[22] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[23] J. Bergstra and Y. Bengio, “Random Search for Hyper-Parameter Optimization,” *Journal of Machine Learning Research*, vol. 13, no. 10, pp. 281–305, 2012, [Online]. Available: http://jmlr.org/papers/v13/bergstra12a.html

[24] W. S. Bhagawan, W. Ekasari, and M. Agil, “Ethnopharmacology of medicinal plants used by the Tenggerese community in Bromo Tengger Semeru National Park, Indonesia,” *Biodiversitas*, vol. 24, no. 10, Nov. 2023, doi: 10.13057/biodiv/d241028.

[25] W. S. Bhagawan, A. Suproborini, D. L. P. Putri, A. Nurfatma, and R. T. Putra, “Ethnomedicinal study, phytochemical characterization, and pharmacological confirmation of selected medicinal plant on the northern slope of Mount Wilis, East Java, Indonesia,” *Biodiversitas*, vol. 23, no. 8, Aug. 2022, doi: 10.13057/biodiv/d230855.

[26] W. S. Bhagawan, W. Ekasari, and M. Agil, “Ethnobotanical survey of herbal steam baths among the Tenggerese community in Bromo Tengger Semeru National Park, Indonesia,” *IOP Conf Ser Earth Environ Sci*, vol. 1352, no. 1, p. 012103, May 2024, doi: 10.1088/1755-1315/1352/1/012103.

[27] Y. Yiblet, “Ethnobotanical study of medicinal plants used to manage human ailments in Lay Gaint District, South Gondar Zone, Amhara Region, Northwestern Ethiopia,” *Heliyon*, vol. 10, no. 15, p. e35277, 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e35277.

[28] B. Paneru, B. Thapa, and B. Paneru, “Leveraging AI in ayurvedic agriculture: A RAG chatbot for comprehensive medicinal plant insights using hybrid deep learning approaches,” *Telematics and Informatics Reports*, vol. 16, p. 100181, 2024, doi: https://doi.org/10.1016/j.teler.2024.100181.

[29] L. Breiman, “Random Forests,” *Mach Learn*, vol. 45, pp. 5–32, 2001, doi: 10.1007/978-3-030-62008-0_35.

[30] D. Shanthini, M. Shanthi, and M. C. Bhuvaneswari, “A Comparative Study of SVM Kernel Functions Based on Polynomial Coefficients and V-Transform Coefficients,” *International Journal Of Engineering And Computer Science*, vol. 6, no. 6, pp. 2319–7242, 2017, doi: 10.18535/ijecs/v6i3.65.

[31] D.- Andriansyah and Eka Wulansari Fridayanthie, “Optimization of Support Vector Machine and XGBoost Methods Using Feature Selection to Improve Classification Performance,” *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 6, no. 2, pp. 484–493, Jan. 2023, doi: 10.31289/jite.v6i2.8373.

[32] R. Azadnia, F. Noei-Khodabadi, A. Moloudzadeh, A. Jahanbakhshi, and M. Omid, “Medicinal and poisonous plants classification from visual characteristics of leaves using computer vision and deep neural networks,” *Ecol Inform*, vol. 82, p. 102683, 2024, doi: https://doi.org/10.1016/j.ecoinf.2024.102683.

[33] M. A. Kiflie, D. P. Sharma, and M. A. Haile, “Deep learning for Ethiopian indigenous medicinal plant species identification and classification,” *J Ayurveda Integr Med*, vol. 15, no. 6, p. 100987, 2024, doi: https://doi.org/10.1016/j.jaim.2024.100987.