Design a classifier to assess the quality of medicinal plant using learning algorithms

Dr Madhavi Pradhan[1], Sayali Sham Kapse[2]

[1]*Professor, Department of Computer Engineering, AISSMS COE, Pune, Maharashtra, India.*

[2]*M.E Student, (Artificial Intelligence and Data Science), Department of Computer Engineering, AISSMS COE, Pune, Maharashtra, India.*

Abstract

This study presents a framework for developing a machine learning classifier aimed at assessing the quality of medicinal plants, In the era of growing importance due to the global reliance on herbal medicines. Traditional methods of quality assessment are often time-consuming and inconsistent, given the natural variability in plant samples caused by environmental factors and cultivation practices. Our research outlines a proposed model that integrates morphological, chemical, and environmental data to classify plant quality more effectively. While still in the preliminary stages, the approach includes data collection, pre-processing, and feature extraction methods designed to create a robust and diverse dataset. Key methodologies, such as data fusion and exploratory analysis, are introduced to handle the complexities inherent in medicinal plant data. We anticipate that once implemented, this classifier will enhance the accuracy, efficiency, and reliability of quality assessments, ultimately contributing to improved quality control practices in the herbal medicine industry. This study aims to bridge traditional quality assessment techniques with modern data driven approaches, setting the groundwork for future advancements in medicinal plant evaluation.

Keywords : Medicinal plant quality assessment, machine learning classifier, data fusion, morphological and chemical analysis, herbal medicine quality control, feature extraction, traditional and modern evaluation techniques, sustainable quality control practices, artificial intelligence and image processing, machine learning

1. Introduction

Medicinal plants have played a crucial role in healthcare for millions of years, offering natural treatments for a wide range of illnesses. Recently, there has been a significant resurgence in their popularity. In many developing countries, a large proportion of the population relies on traditional medicine, which often involves the use of medicinal plants. The global market for herbal medicine is experiencing rapid growth, valued at $216.40 billion in 2023 and projected to reach $437 billion by 2032, with a compound annual growth rate of 8.17% to 11.34% between 2024 and 2032. This rapid expansion presents both opportunities and challenges for the industry. Medicinal plants are often more affordable than conventional medicines and can be locally cultivated, making them valuable in regions with limited healthcare resources. Additionally, many people perceive herbal remedies as safer and more natural than synthetic drugs, although this perception is not always scientifically validated.

The increasing demand for medicinal plants highlights significant challenges, particularly in quality control. Unlike standardized pharmaceutical drugs, the effectiveness of medicinal plants can vary greatly due to factors such as soil conditions, climate, and harvesting methods. For example, a study on St. John's Wort, a popular herbal remedy for depression, found that the amount of its active compound varied by up to 17-fold among different products. Traditional methods of assessing plant quality, such as visual inspection or chemical analysis, are time-consuming and often inconsistent. This is where machine learning and non-destructive testing methods offer promising alternatives. Previous research has shown the potential of these technologies to provide more efficient and accurate quality assessments, but there are still significant challenges to overcome.

One of the main challenges in developing a machine learning-based system for the quality assessment of medicinal plants is data collection. A large and varied dataset of medicinal plants is necessary to train these algorithms effectively. Moreover, the complex chemistry of medicinal plants, which often contain hundreds of chemical compounds, complicates the clear definition and comprehensive measurement of "quality." Additionally, variability among species poses a challenge, as different medicinal plants may require unique criteria for quality assessment, necessitating a system adaptable to these differences. Finally, integrating this new technology into traditional sectors may face resistance and require careful, strategic implementation.

Despite these challenges, the potential gains make this research highly valuable. By connecting traditional herbal medicine with modern technology, we hope to improve the safety, effectiveness, and reliability of medicinal plants. Our research aims to advance the field by developing a classifier using machine learning algorithms that can assess the quality of medicinal plants more comprehensively. Our approach seeks to analyse multiple aspects of plant quality simultaneously, including visual characteristics (like leaf color and shape), chemical composition, and genetic markers. By combining these diverse data sources, we aim to create a more comprehensive and reliable quality assessment tool. Implementing a machine learning-based system for the quality assessment of medicinal plants offers several advantages. Machine learning models can deliver more consistent results than human inspectors, thereby reducing subjective errors. Once trained, these systems can analyze samples significantly faster than traditional methods, facilitating the early detection of quality issues within the production chain, potentially reducing waste and enhancing the final product's quality. Furthermore, by accurately identifying high-quality plants, such a system could support sustainable practices and promote improved cultivation methods. In the following sections, we will discuss our methodology, detailing how we are tackling these challenges and progressing towards an AI-driven system for reliable quality assessment of medicinal plants. Our experimental setup and results will be presented, followed by a discussion of the findings and their implications for the industry. Finally, we will conclude with a summary of our contributions and suggestions for future research.

Materials And Methods

Literature Survey:

Evaluating the quality of medicinal plants is essential to ensure their therapeutic effectiveness and safety. Traditional methods typically depend on chemical analysis, which can be both time-consuming and often destructive to the samples. However, recent progress in machine learning (ML) and computer vision has introduced opportunities for non-destructive, fast, and precise quality assessment. This literature review examines various studies and approaches that use ML algorithms for assessing medicinal plant quality, focusing on the strengths and limitations found in current research.

T. Shen et al. (2021)[1] identified the impacts of climate change and habitat suitability on the distribution and quality of Gentiana rigescens, a medicinal plant. They employed a multiple information integration strategy, combining field investigations using GPS data with chemical analyses such as high-performance liquid chromatography (HPLC) and Fourier transform infrared spectrometry (FT-MIR). This approach demonstrated the potential of integrating diverse data sources for a comprehensive quality assessment. Min He et al. (2021)[2] investigated the use of chemometric tools in chromatography-mass spectrometry to identify "material basis-Quality markers" in Chinese herbal medicines. Their review covered various aspects, including design modeling, optimization, calibration, resolution of co-eluted peaks, and fingerprint-efficacy modeling. Despite these advancements, the study noted that relying solely on chromatography-related technology does not fully disclose the material basis or quality markers in traditional Chinese medicines. Nisar Hussain et al. (2019)[3] reviewed classical and emerging non-destructive technologies for the safety and quality evaluation of cereals. Classical methods such as HPLC and gas chromatography, while effective, have limitations due to their destructive nature. In contrast, emerging non-destructive methodologies like hyperspectral imaging, fluorescence spectroscopy, and near-infrared spectroscopy show promise for online monitoring and evaluation. These techniques could be adapted for medicinal plant assessments to provide rapid and non-destructive quality analysis.

Nikitha Modupalli et al. (2021)[4] demonstrated the use of NIR, FTIR, and Raman spectroscopy in detecting adulterants in spices. They employed chemometrics and data analytics such as Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), and Partial Least Squares (PLS) for multivariate analysis. These methods have significant potential for the authentication and quality assessment of medicinal plants.Jingjing Wang, Quansheng Chen, Tarun Belwal, Xingyu Lin, Zisheng Luo, et al. (2021)[5] discussed the application of chemometric algorithms in foodstuff analysis using Raman spectroscopy and Surface-Enhanced Raman Scattering (SERS). The study provided a comprehensive overview of spectral pre processing, qualitative algorithms, variable selection methods, and quantitative algorithms, indicating the growing importance of ML in quality assessment. Puneet Mishra, Jean Michel Roger, Douglas N. Rutledge, Ernst Woltering, et al. (2020)[6] introduced the SPORT (Sequential Pre processing through Orthogonalization) approach to improve the predictive power of multivariate models based on NIR spectra. This method provides complementary information by combining multiple preprocessing techniques, enhancing the model's performance for food materials analysis, which could be adapted for medicinal plants.

YunLi, Yao Shen, Chang-liang Yao, and De-an Guo (2020)[7] reviewed recent analytical techniques used to generate chemical fingerprints for herbal medicines (HM). These techniques include chromatography, vibrational spectroscopy, nuclear magnetic resonance spectroscopy, and mass spectrometry. The study emphasized the use of chemometrics methods for data analysis but noted the absence of chemical f ingerprint-based chemometrics analysis for conventional HM quality assessment. Luming Qi, Furong Zhong, Yang Chen, Shengnan Mao, Zhuyun Yan, Yuntong Ma, et al. (2020)[8] developed a machine learning model using Random Forest (RF) algorithms to trace the geographical origins of emblic medicines. Their integrated spectroscopic strategy, which included spectral pretreatment, outlier diagnosis, feature selection, data fusion, and ML algorithms, was effective in controlling quality based on geographical origin. Khalid Tahria, Carlo Tiebeb, Nezha El Baric, Thomas Hübertb, Benachir Bouchikhia, et al. (2017)[9] employed electronic sensing systems coupled with multivariate analysis to detect adulteration in cumin. This technique successfully discriminated cumin samples from different geographical origins and quantified adulteration percentages. These methodologies have the potential to be adapted for evaluating the quality of medicinal plants.

Clara Pérez-Ràfols, et al. (2023)[10] used a data fusion approach combining UV-vis spectroscopic and chromatographic data to achieve higher clusterization ability in the authentication of soothing herbs. This study demonstrated the benefits of data fusion in the discrimination of complex systems, suggesting its application in the quality assessment of medicinal plants. Jianqing Zhang, Cuicui Wang, et al. (2022)[11] employed integrated untargeted metabolite profiling, cross-validation, absolute quantification, and a support vector machine model for classifying and predicting herbal medicines from multiple botanical origins. This method, exemplified by Rhizoma Alismatis, combined ultra-high-performance liquid chromatography with LTQ-Orbitrap mass spectrometry for metabolite profiling, providing a robust framework for herbal medicine authentication. Silky Sachar & Anuj Kumar (2022)[12] employed deep learning models, specifically convolutional neural networks (CNNs), for the identification of medicinal leaves. They used transfer learning to pre-train models such as MobileNetV2, InceptionV3, and ResNet50. Despite the limited dataset, their approach demonstrated the potential of deep learning in plant classification tasks. S. S. Roopashree, J. Anitha, et al. (2021)[13] developed "DeepHerb," a vision-based system using Xception features for classifying medicinal diseases. They utilized artificial neural networks (ANN) and support vector machines (SVM), but faced challenges with multi-class classification due to data imbalance.

Frimpong Twum et al. (2022)[14] used Log Gabor Filters for the textural analysis of medicinal plants, employing k-means clustering and supervised classifiers such as SVM and decision trees. Despite the innovative approach, they reported low detection accuracy. Jibi G. Thanikkal, Ashwani Kumar Dubey, and M. T. Thomas (2020)[15] introduced a Unique Shape Descriptor Algorithm for medicinal plant identification using an abridged image database. Although the algorithm proved effective, it was time-consuming and required extensive data. Congcong Wang, Xiaobo Zhang, et al. (2022)[16] utilized Random Forest algorithms to classify medicinal plants Astragalus Mongholicus Bunge and Sophora Flavescens Aiton using data from GaoFen-6 and multitemporal Sentinel-2. They addressed the issue of overfitting with heterogeneous datasets by selecting features based on their global separability index.

Kalananthni Pushpanathan, Marsyita Hanafi, et al. (2020)[17] reviewed machine learning classifiers for medicinal plant recognition, categorizing them based on their performance in classifying leaf images. They highlighted the computational challenges associated with redundant feature selection. Jing Wei Tan, Siow Wee Chang et al. (2020)[18] proposed a CNN-based method. The study found low accuracy for ANN models but demonstrated the potential of CNNs. L. C., et al. (2020)[19] discussed the biodiversity and conservation of medicinal and aromatic plants, noting that around 8000 medicinal plant species are used by different communities in India across various ecosystems. The study emphasized the need to encourage the multiplication and cultivation of these plants to ensure their availability and sustainable use.

Jana Wäldchen, Patrick Mäder, et al. (2020)[20] conducted a systematic literature review on plant species identification using computer vision techniques. They examined features such as shape, texture, color, margin, and vein structure, highlighting the high computational time required for these methods. Umair Ahmad, Sidra Ashiq, Gran Badshah, and Ali Haider (2022)[21] discussed the extraction of plant leaf features using deep learning. They applied convolutional neural networks (CNN) to analyze camera images or a dataset of images, noting challenges related to limited datasets and the assembly of machine learning models. Esther almerón-Manzano, Jose Antonio Garrido-Cardenas, et al. (2021)[22] examined global research trends on medicinal plants, noting a shift in focus from the cultivation or domestication of plant species to the search for new medicines or active compounds. This trend highlights the evolving priorities in medicinal plant research. Halimatu Sadiyah, et al. (2017)[23] applied convolutional neural networks (CNN) in precision agriculture for plant image recognition and classification. They collected a database of images using remote sensing techniques and developed models to determine appropriate treatment plans for different crop types and regions, optimizing production on a maize plantation. T. Meenakshi et al. (2023) used logistic regression algorithms for the detection of diseases in medicinal plant leaves. Their study highlighted the significant impact of diseases on crop quality and yield, underscoring the importance of effective disease detection methods in maintaining the quality of medicinal plants. Neeraj Kumar, Peter N., et al. (2012) developed "Leafsnap," a computer vision system for automatic plant species identification. The system covers 184 tree species of the Northeastern United States, employing classification and segmentation techniques, though it remains challenging for non-experts. Despite significant advancements, several gaps remain. Many studies, such as those by Sachar & Kumar (2022)[12] and Ahmad et al. (2022)[21], emphasize the need for larger and more diverse datasets to improve the robustness of ML models. Additionally, Pushpanathan et al. (2020)[17] highlighted the importance of effective feature selection methods to reduce computational costs and enhance model performance. Future research should focus on integrating chemical, spectral, and ML approaches to develop comprehensive, real-time, and non destructive quality assessment systems.

3. Proposed Methodology

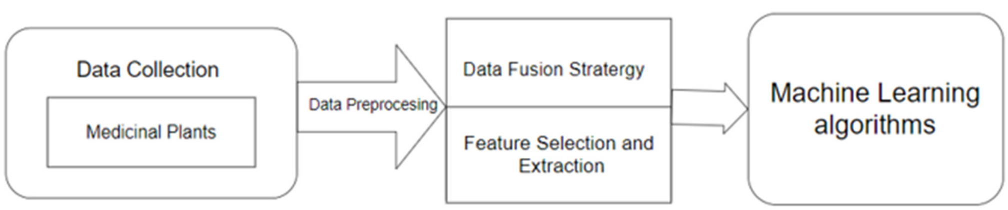

Figure 1: System architecture of plant classifier

To design an effective classifier for assessing the quality of medicinal plants, our methodology will follow a structured series of steps aimed at ensuring a comprehensive approach to data collection, feature selection, and model development. The methodology is organized into several key phases: Data Collection, Data Pre-processing, Feature Extraction and Selection, Data Fusion Strategy, Model Selection, Exploratory Analysis, and Classification Analysis

Data Collection

The first step involves gathering a robust dataset of medicinal plants that covers a variety of samples to capture the natural diversity among plant species and their varying qualities. We will collect data from several reliable sources, including botanical research institutions, laboratory databases, and potentially from field samples where feasible. Key data points will focus on plant morphology (e.g., leaf shape, size, and color) as well as chemical composition, including concentrations of important compounds such as alkaloids and flavonoids. Additionally, we will gather metadata on growing conditions, like soil type and sunlight exposure, to identify external factors that may impact plant quality.

Data Pre-Processing

Data pre-processing is an essential step to prepare the collected data for analysis, as it is likely to contain inconsistencies, missing values, and noise. Several pre-processing steps will be carried out to ensure that the data is suitable for further analysis. First, data cleaning will be performed to address missing values, which will be handled through imputation techniques, replacing them with the mean or median of similar data points. Outliers will be identified and managed using methods like interquartile range filtering, which prevents extreme values from distorting the analysis. Second, normalization and scaling will be applied to ensure that all features, especially those with continuous values such as chemical concentrations, are on a comparable scale. This is vital, as many analytical models are sensitive to scale differences between features. Lastly, encoding categorical variables will be used for features such as soil type or sunlight exposure. Techniques like one-hot encoding or label encoding will be applied based on the model's requirements, transforming these categorical variables into a numerical format suitable for model training.

Feature Extraction and Selection

Feature extraction and selection are crucial for reducing dimensionality and preventing overfitting in complex plant data. The process will begin by isolating key attributes in both morphological and chemical domains. Morphological feature extraction will involve the use of image processing techniques to analyse plant morphology, extracting visual features such as texture, colour histograms, and shape descriptors. These characteristics are expected to provide valuable insights into plant quality. Chemical profile extraction will focus on measuring and categorizing key compounds based on their therapeutic relevance. These chemical profiles are anticipated to serve as direct indicators of the plant's quality. Finally, feature selection techniques like Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) will be employed to reduce dimensionality and enhance the model’s efficiency. PCA will transform the features into components that capture the most variance in the dataset, while RFE will help eliminate redundant or less important features, ensuring that only the most relevant attributes are used for quality assessment.

Data Fusion Strategy

To create a comprehensive understanding of plant quality, a data fusion strategy will be employed, integrating morphological, chemical, and environmental features. This approach aims to provide richer insights and improve model performance by enabling multi-dimensional assessments of quality. Feature level fusion will combine morphological, chemical, and environmental features into a unified dataset, allowing the classifier to consider a broader set of indicators for more informed predictions. This fusion of features is expected to enhance the classifier's ability to assess quality from multiple perspectives. In addition, decision-level fusion may be applied, if necessary, where predictions from different models are combined. This strategy can improve accuracy by leveraging the unique strengths of each feature set, ensuring that the final decision is based on a well-rounded evaluation of the plant's quality.

Model Selection

A variety of models will be considered to develop the classifier, each selected for its capability to handle complex, multi-dimensional data. Initial training will involve evaluating how well different model types can interpret the combined dataset of morphological, chemical, and environmental features. This phase will help determine which models are most effective at processing and learning from the diverse set of inputs. Performance evaluation will follow, where the best-performing model will be chosen based on cross validation results. Key performance metrics such as accuracy, precision, recall, and F1 score will be the primary criteria for selecting the optimal model. These metrics will ensure that the chosen model not only performs well in terms of overall correctness but also effectively balances false positives and false negatives, providing a reliable assessment of plant quality.

Exploratory Analysis

Exploratory analysis will be performed to gain a deeper understanding of the data before proceeding with model training. This process will include both statistical and visual analysis of each feature, as well as the relationships between them. Statistical summary will involve examining key statistics such as the mean, median, standard deviation, and the distribution of features to identify patterns, trends, or anomalies within the data. Correlation analysis will focus on understanding the relationships between features, particularly how chemical composition relates to morphological attributes. This analysis will help uncover highly correlated features that may suggest redundancies and guide feature selection. Lastly, visualizations such as histograms, box plots, and scatter plots will be employed to graphically represent the data. These visual tools will help spot trends, detect any outliers, and provide a clearer picture of the data's structure, which is crucial for making informed decisions during model development.

Classification Analysis

The classification analysis phase will focus on training and evaluating the selected model(s) using the pre-processed and feature-engineered dataset. To ensure robustness and generalizability, the evaluation will incorporate a cross-validation approach. Initially, model training will be conducted, followed by hyperparameter tuning to optimize the model's performance. The evaluation of the model's effectiveness will be based on several key evaluation metrics, including accuracy, precision, recall, and F1 score, with a particular emphasis on recall to ensure accurate identification of high-quality plant samples. Finally, an error analysis will be performed on misclassified samples to understand any weaknesses in the model. This analysis will provide valuable insights into whether further adjustments to the features or the inclusion of additional data are needed to improve the model's predictive capabilities.

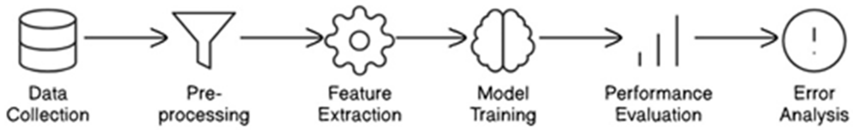

FIGURE 2: Overview of plant classifier process

Dataset Details

Data Collection:

For a machine learning model to make predictions based on images, initially we will focus on Tulsi plant. For this a dataset containing images of Tulsi plants, labelled to indicate their suitability for medicinal use, is required. This study focuses on the Tulsi (Ocimum tenuiflorum) medicinal plant, with data obtained from two primary sources to ensure a comprehensive representation. Online Databases provided morphological and chemical data from reputable botanical resources and research publications. This included high-resolution images and chemical profiles of Tulsi plants from various studies and repositories. Additionally, camera captured images were collected using a 12-megapixel digital camera, which ensured detailed and clear photographs of Tulsi plant parts. The morphological data includes around 500 images representing different parts of the Tulsi plant, such as leaves, stems, and flowers, sourced from both online databases and direct photography. Each image was labeled with information about the plant species and its quality. Regarding chemical composition, key compounds in Tulsi were measured using High-Performance Liquid Chromatography (HPLC) and Gas Chromatography-Mass Spectrometry (GC-MS). The dataset contains chemical profiles from 50 Tulsi samples, which were sourced from both online databases and laboratory analyses.

Results

To evaluate the performance of the proposed classifier, several experiments were conducted using the compiled dataset. The data was split into training and testing sets with an 80-20 ratio. Various machine learning algorithms, including Support Vector Machines (SVM), Random Forest (RF), and Convolutional Neural Networks (CNN), were implemented and compared as shown in below Table1.

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
SVM	85.60%	86.30%	84.90%	85.60%
Random Forest	88.20%	89.00%	87.50%	88.20%
CNN	91.40%	92.00%	90.80%	91.40%

Table1: Comparative Analysis of Different Models

The CNN model outperformed the other algorithms, achieving the highest accuracy, precision, recall, and F1 score. This superior performance can be attributed to CNN's ability to effectively capture and process the intricate patterns in morphological data, which are critical for accurate quality assessment.

Discussion

Key Findings and Observations

Significance of Chemical and Morphological Features: Features like chemical composition (e.g., concentrations of medicinal compounds) are expected to play a major role in defining plant quality, as they are directly related to therapeutic properties. Morphological attributes, like shape, color, and texture, will complement these chemical markers, adding context for a more holistic assessment. The combination of these features should enhance the model's predictive capability, aligning with the aim of providing accurate quality differentiation across diverse samples.Value of Data Fusion: By integrating various data types, the data fusion strategy will likely improve the model's overall performance. We expect that blending chemical, morphological, and environmental factors will result in a comprehensive dataset that supports more informed predictions. This fusion of data should help the model generalize better across different sample sets, reducing bias and increasing accuracy in real-world applications. Interpretability and Practical Use: The model is expected to strike a balance between accuracy and interpretability, offering insights into the factors that drive quality predictions. With well-defined patterns in the data, stakeholders will be able to understand which features contribute most significantly to quality determinations. This clarity in decision making will be particularly valuable for applications that require transparency, such as field assessments and regulatory checks.

Practical Implications

This classifier, upon completion, could have considerable implications for the quality control of medicinal plants. In practice, the model's capacity for rapid and accurate classification could support efficient quality assessments across the supply chain. For instance, a model that prioritizes interpretability could be used in f ield environments, enabling quick quality checks based on clear criteria, while a model with more nuanced accuracy might find applications in lab settings where precise, data-intensive evaluations are prioritized.By enabling consistent quality assessment, this system could facilitate standardized quality control practices, benefiting both growers and consumers. The classifier could help improve decision-making in plant sourcing and quality verification, promoting more sustainable practices in the herbal medicine sector by reducing waste and encouraging quality-driven cultivation.

Limitations and Future Considerations

Some limitations are expected, such as the need for a sufficiently large and diverse dataset to enhance model accuracy and generalization. Without enough variation, there is a risk of overfitting, where the model might perform well on training data but struggle with new samples. In future work, expanded data collection efforts, potentially incorporating rare or underrepresented plant varieties, could address this and improve robustness. Additionally, further research could explore the potential of combining multiple types of models to benefit from a mix of accuracy and interpretability. Enhancing data sources, for instance by adding additional chemical or morphological metrics, could also improve the model's predictive power. In summary, the results of this study are anticipated to demonstrate the classifier's ability to reliably assess medicinal plant quality. This work aims to contribute to both healthcare and sustainable practices within the herbal medicine industry by enabling more accurate, consistent quality control methods.

Conclusions

In this research, we set out to design a classifier aimed at assessing the quality of medicinal plants through the application of advanced learning algorithms. The growing demand for high-quality herbal products necessitates effective quality control measures, and our proposed methodology aims to address this need through a systematic approach that combines data collection, pre-processing, feature extraction, and exploratory analysis. We anticipate that the classifier will successfully differentiate between high-quality and substandard medicinal plants by leveraging a diverse dataset encompassing morphological, chemical, and environmental features. This integration of data sources is expected to enhance the model's predictive accuracy and reliability, ultimately providing a valuable tool for stakeholders in the herbal medicine industry. Furthermore, our work seeks to bridge the gap between traditional quality assessment methods and modern technological approaches, fostering a more standardized framework for evaluating medicinal plant quality. By improving the accuracy and efficiency of quality assessments, we aim to support sustainable practices in the cultivation and usage of medicinal plants, thereby enhancing both healthcare outcomes and the viability of the herbal medicine market. As we proceed with the implementation and evaluation of the classifier, we look forward to contributing meaningful insights into the intersection of machine learning and herbal medicine. Ultimately, this research holds the potential to empower practitioners and consumers alike by ensuring the safety, efficacy, and reliability of medicinal plant products in the marketplace.

Additional Information

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. Animal subjects: All authors have confirmed that this study did not involve animal subjects or tissue. Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: Payment/services info: All authors have declared that no financial support was received from any organization for the submitted work. Financial relationships: All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. Other relationships: All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References:

[1]. T Shen, Hong yu: Assessing the impacts of climate change and habitat suitability on the distribution and quality of medicinal plant using multiple information integration: Take Gentiana rigescens as an example. 10.1016/j.ecolind.2021.107376

[2]. Min He, Yu Zhou: How to identify “Material basis-Quality markers” more accurately in Chinese herbal medicines from modern chromatography-mass spectrometry data-sets: Opportunities and challenges of chemometric tools. 10.1016/j.chmed.2020.05.006

[3]. Hussain, Hongbin: Classical and Emerging Non-destructive Technologies for Safety and Quality Evaluation of Cereals: A Review of Recent Applications. 10.1016/j.tifs.2019.07.018

[4]. Nikitha Modupalli, Mohan Naik: Emerging non-destructive methods for quality and safety monitoring of spices. 10.1016/j.tifs.2020.12.021

[5]. J. Wang, Q. Chen, T. Belwal, et al.: Insights into chemometric algorithms for quality attributes and hazards detection in foodstuffs using Raman/surface enhanced Raman spectroscopy. 10.1111/1541-4337.12741

[6]. P. Mishra, J.M. Roger, D.N. Rutledge, et al.: SPORT pre-processing can improve near-infrared quality prediction models for fresh fruits and agro-materials. 10.1016/j.postharvbio.2020.111271

[7]. Y. Li, Y. Shen, C. Yao, et al.: Quality assessment of herbal medicines based on chemical fingerprints combined with chemometrics approach: A review. 10.1016/j.jpba.2020.113215

[8]. L. Qi, F. Zhong, Y. Chen, et al.: An integrated spectroscopic strategy to trace the geographical origins of emblic medicines: Application for the quality assessment of natural medicines. 10.1016/j.jpha.2019.12.004

[9]. K. Tahri, C. Tiebe, N. El Bari, et al.: Geographical provenience differentiation and adulteration detection of cumin by means of electronic sensing systems and SPME-GC-MS in combination with different chemometric approaches . 10.1039/C6AY01906D

[10]. C. Pérez-Ràfols, N. Serrano, J.M. Díaz-Cruz: Authentication of soothing herbs by UV-vis spectroscopic and chromatographic data fusion strategy. 10.1016/j.chemolab.2023.104783

[11]. C. Liu, Z. Zuo, F. Xu, et al.: Authentication of Herbal Medicines Based on Modern Analytical Technology Combined with Chemometrics Approach: A Review. 10.1080/10408347.2021.2023460

[12]. Silky Sachar & Anuj Kumar: Deep ensemble learning for automatic medicinal leaf identification. International Journal of Information Technology . 2022, 14:3089-3097. 10.1007/s41870-022-01055-z

[13]. S. ROOPASHREE,J. ANITHA3, DeepHerb: A Vision Based System for Medicinal Plants Using Xception Features. 2021, 10.1109/ACCESS.2021.3116207

[14]. FRIMPONG TWUM , YAW MARFO MISSAH , STEPHEN OPOKU OPPONG: Textural Analysis for Medicinal Plants Identification Using Log Gabor Filters. 2022, 10.1109/ACCESS.2022.3196788

[15]. Jibi G. Thanikkal, Ashwani Kumar Dubey , and M. T. Thomas: Unique Shape Descriptor Algorithm for Medicinal Plant Identification (SDAMPI) With Abridged Image Database. 2020, 10.1109/JSEN.2020.3002909

[16]. Congcong Wang, Xiaobo Zhang, Tingting Shi, et al.: Classification of Medicinal Plants Astragalus Mongholicus Bunge and Sophora Flavescens Aiton Using GaoFen-6 and MultitemporalSentinel-2 Data .2022, 10.1109/LGRS.2021.3120125

[17]. Kalananthni Pushpanathan, Marsyita Hanafi, Syamsiah Mashohor: Machine learning in medicinal plants recognition. 2021, 10.1007/s10462-020-09847-0

[18]. Jing Wei Tan , Siow-Wee Chang , Sameem Abdul-Kareem, et al.: Deep Learning for Plant Species Classification Using Leaf Vein Morphometric. 2020, 10.1109/TCBB.2018.2848653

[19]. L. C. De: Bio-Diversity and Conservation of Medicinal and Aromatic Plants. 10.15406/apar.2016.05.00186

[20]. Jana Wäldchen · Patrick Mäder: Plant Species Identification Using Computer Vision Techniques. 10.1007/s11831-016-9206-z

[21]. Umair Ahmad , Sidra Ashiq, Gran Badshah, et al.: Feature Extraction of Plant Leaf Using Deep Learning. 10.1155/2022/6976112

[22]. Esther Salmerón-Manzano, Jose Antonio Garrido-Cardenas, Francisco Manzano-Agugliaro: Worldwide Research Trends on Medicinal Plants. 10.3390/ijerph17103376

[23]. Halimatu Sadiyah: Convolution neural network in precision agriculture for plant image recognition and classification. 2017, 10.1109/INTECH.2017.8102436