International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Development of Leaf Extract Spectral Database for Plant Species Classification Using Liquid Light Guide Spectroscopy

Gaju S. Chavan[1] Sonali B. Kulkarni[2] Simran N. Maniyar[3]

[1,2,3]Department of Computer Science and Information Technology
Dr. Babasaheb Ambedkar Marathwada University, Chhatrapati Sambhajinagar,

Abstract

Plant species identification is crucial in agriculture, conservation of biodiversity, forestry, medicinal plant authentication and precision farming. The traditional identification using morphological features is often time consuming, requires expert knowledge and has limitations due to environmental changes and seasonal variations. In this study, developed a standard library of leave extracts spectra for automatic plant species identification using transmission spectroscopy. Fresh, healthy leaves were used to prepare leaf extracts which were developed spectral library using the ASD FieldSpec4 spectroradiometer with Liquid Light Guide (LLG) over the spectral range of 350–2500 nm in the visible (VIS), near infrared (NIR), and short-wave infrared (SWIR) spectral regions. Several biological samples were taken for each plant species and 10 spectra were recorded for each sample after calibration with dark and blank samples to increase the reliability of the measurements. The recorded spectra were averaged and prepared in a hierarchical spectra database based on species, biological material and repeated spectrum acquisition. Proposed transmission spectroscopy is a different method to conventional reflectance spectroscopy, which is able to measure biochemical signatures related to pigments, water, carbohydrates, proteins and other metabolites in leaf extracts. This spectral database will be use as a credible basis for the spectral pre-processing, machine learning and plant species classification based on artificial intelligence. This effort provides a standardized dataset that facilitates the development of strong, repeatable and non-destructive plant identification systems for future use in the agricultural and environmental domains.

Keywords: Liquid Light Guide, Spectral Transmission, biodiversity, Plant Species, Spectroscopy.

1. Introduction

Accurate identification of plant species is crucial in agriculture, forestry, biodiversity conservation, authentication of medicinal plants, and precision farming. Traditionally, the main ways of identifying plants are based on morphology, such as leaf shape, flower structure, stem morphology and taxonomic knowledge [1]. These methods are widely used but are often time-consuming, labour-intensive and highly dependent on expert knowledge. Furthermore, morphological traits are sensitive to modifications influenced by environmental factors, age, seasonal variations or disease; therefore, visual identification is not always reliable.

The identification of plant species has become faster and non-destructive due to recent progress in spectroscopy and artificial intelligence. Spectroscopic techniques rely on the measurement of the interaction of electromagnetic radiation with biological materials. This allows the extraction of unique spectral signatures related to the biochemical composition of plant tissues [2]. These spectral signatures contain useful information about pigments, water content, carbohydrates, proteins, phenolic compounds, chlorophyll, and other metabolites, making spectroscopy a successful tool for plant characterization.

The ASD FieldSpec 4 spectroradiometer has been widely used for vegetation analysis among the available spectroscopic instruments due to its high spectral resolution and wide wavelength coverage from 350 to 2500 nm including the visible (VIS), near-infrared (NIR) and short-wave infrared (SWIR) regions. Prior work has mainly addressed the measurement of spectral reflectance directly from intact leaves or plant canopies. Reflectance spectroscopy has shown promising classification performance, but it mostly measures surface optical properties and might not fully exploit the biochemical information associated with plant tissues [3].

An alternative approach of liquid spectroscopy has been emerged for determining plant biochemical properties. By extracting leaf metabolites into a liquid medium and measuring transmission spectra with a Liquid Light Guide (LLG), spectral information which is highly related to the internal chemical composition of the plant can be acquired. The leaf extract transmission spectra possess unique spectral patterns which can be used as chemical fingerprints for the discrimination of plant species [4]. Nevertheless, public databases of spectra obtained with Liquid Light Guide transmission spectroscopy are still scarce despite the growing use of spectrum analysis in plant science.

Development of the standard spectral database is crucial to developing machine learning algorithms that can automatically classify plants into their respective species. Such a database should contain a large number of biological specimens and also be based on repeat spectral readings, which will help generalize the models and allow development of intelligent decision-making systems in agriculture.

This study was conducted by create a leaf database of their extract spectra with an ASD FieldSpec 4 equipped with a Liquid Light Guide (LLG) Spectroscopy. Leaf extracts were prepared from leaf samples taken from various plant species and spectra were measured for transmission within the range 350 – 2500 nm. Multiple scans were taken for each biological sample to enhance measurement accuracy.

2. Related Work

The identification of plant species has become a topic of great interest in recent years for various purposes such as agriculture, conservation of biodiversity, forestry, authentication of medicinal plants and precision farming [5], [6]. Traditional taxonomic systematics uses the following characteristics: Leaf shape, flower morphology, stem structure and fruit characteristics. These methods are accepted, but they are time consuming, sensitive to environmental conditions, seasonal changes, and plant growth stages [7]. To this end, researchers have investigated spectroscopic techniques that can be used in conjunction with artificial intelligence (AI) to carry out the rapid and objective identification of plant species, which can be automated [8].

Hyperspectral spectroscopy has been shown to be a highly effective method for characterizing biochemical and structural properties of plants. The continuous spectral information collected over hundreds or thousands of wavelengths gives detailed information on various biochemical constituents such as pigments, moisture content, proteins, carbohydrates, lignin, cellulose, and others [9]. Due to its high spectral resolution and wide wavelength range (350 to 2500 nm), the ASD FieldSpec4 series of spectroradiometers has emerged as one of the most popular instruments for vegetation spectroscopy. They are widely used in vegetation monitoring, crop health assessment, disease detection and species discrimination [10].

Leaf reflectance spectroscopy as a means for plant identification has been the subject of several studies. Reflectance spectra taken from leaves in intact plants have been successfully used for reflectance-based classification using machine learning algorithms like Support Vector Machine (SVM), Random Forest (RF), Artificial Neural Networks (ANN) and Deep Learning models [11], [12]. They have proven to have high classification accuracy, but reflectance measurement is mainly the optical property of the surface of the leaves and may not be sufficient to reflect the biochemical composition of the plant tissues.

The major development of recent research has been in the area of spectral preprocessing to enhance classification performance. A number of methods have been applied to try to reduce the spectral noise and to improve informative spectral features such as Savitzky–Golay smoothing, baseline correction, Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV), normalization, derivative spectroscopy and Principal Component Analysis (PCA) [13], [14]. These preprocessing techniques greatly enhance the performance of machine learning models in terms of their robustness and generalizability.

A key component to spectral analysis is the use of machine learning. SVM is known to be effective in high dimensional spectral data and Random forest is robust to noise and offers variable importance estimation [15]. Other classifiers, such as Decision Tree, K-Nearest Neighbor, Logistic Regression, and ensemble learning have been tested on plant species classification with hyperspectral data and are also shown to be successful [16]. With large annotated spectral datasets, deep learning methods such as Convolutional Neural Networks (CNNs) have recently been shown to work well [17].

Although these advances have been made, most published research is based on the use of leaf reflectance spectroscopy or hyperspectral imaging. There have been comparatively few studies that have looked at transmission spectroscopy of extracts, where the biochemical compounds in the leaves are dissolved in a liquid medium [18]. Finally, information is provided from transmission spectroscopy, which directly measures the interaction of light with the dissolved biochemical constituents, not only surface reflectance. This can be used to obtain species-specific chemical signatures which could be used to enhance classification accuracy.

In optical spectroscopy, Liquid Light Guide (LLG) technology has been extensively employed for the transfer of light between light source, sample holder and spectrometer [19]. Its high optical transmitting efficiency, flexibility and wide spectrum response makes it appropriate for laboratory-based transmission measurements. However, its use for the creation of plant spectral databases for automatic plant species classification is still restricted. The majority of the present studies focus on chemical analysis or pigment estimation or quality assessment, while neglecting systematic database construction for machine learning application.

The other major drawback of current research is that there are currently no publicly available, standardized spectral databases of transmission spectra of leaf extracts. The published datasets are mostly based on hyperspectral reflectance images or field measurements of vegetation [20][21]. Lack of well-designed transmission spectral databases hinders the repeatability of research and the building of strong artificial intelligence models for biochemical plant recognition.

This study aims to overcome these limitations by creating a database of leaf extract spectral data with a Liquid Light Guide (LLG) integrated into an ASD FieldSpec 4 spectroradiometer. The method proposed is different from traditional reflectance-based studies, which use transmission spectra of extracts of leaves obtained from 350 to 2500 nm wavelength. The database includes several biological samples and repeated spectral acquisitions for each sample to increase the reliability. In addition, extensive spectral pre-processing and supervised machine learning algorithms are used for automatic species classification of plants. The data set can serve as a starting point for future studies in the areas of transmission spectroscopy, chemical fingerprinting, and plant identification using artificial intelligence.

3. Materials and Methods

3.1 Database

The spectral transmission of vegetation is an approach of collecting data through liquid light guide with ASD fieldspec4 spectroradiometer. These instruments capture the spectral information from the plant at different wavelength across the electromagnetic spectrum. Optical remote sensing utilizes the spectral signature of plants leaf as a method for data collection. The genetic and compound constituents, including leaf-pigments, water-content and dry matter-content, generate a diverse range of absorption characteristics in the spectrum response.

3.1.1 Experimental Setup

Leaf extracts were then tested with an ASD FieldSpec4 spectroradiometer prepared with a Liquid Light Guide (LLG), to obtain transmission spectra. The spectrometer wavelengths from 350 to 2500nm to include the Visible (VIS), Near-Infrared (NIR) and Short-Wave Infrared (SWIR) spectral regions.

The experiment starts with a Liquid Light Guide (LLG) which is attached to a halogen light source. The halogen lamp provides broadband continuous light across the visible (VIS), near infrared (NIR) and short-wave infrared (SWIR) region when operated with the LLG. This light is efficiently transferred to the sample by the LLG, and transfer of heat is minimized. This will avoid over heating of the leaf when taking the spectral measurements, and will ensure a uniform and stable illumination. The spectrometer device was allowed to warm up for approximately 30 minutes before spectral acquisition to ensure stable instrument performance.

The experimental setup consists of:

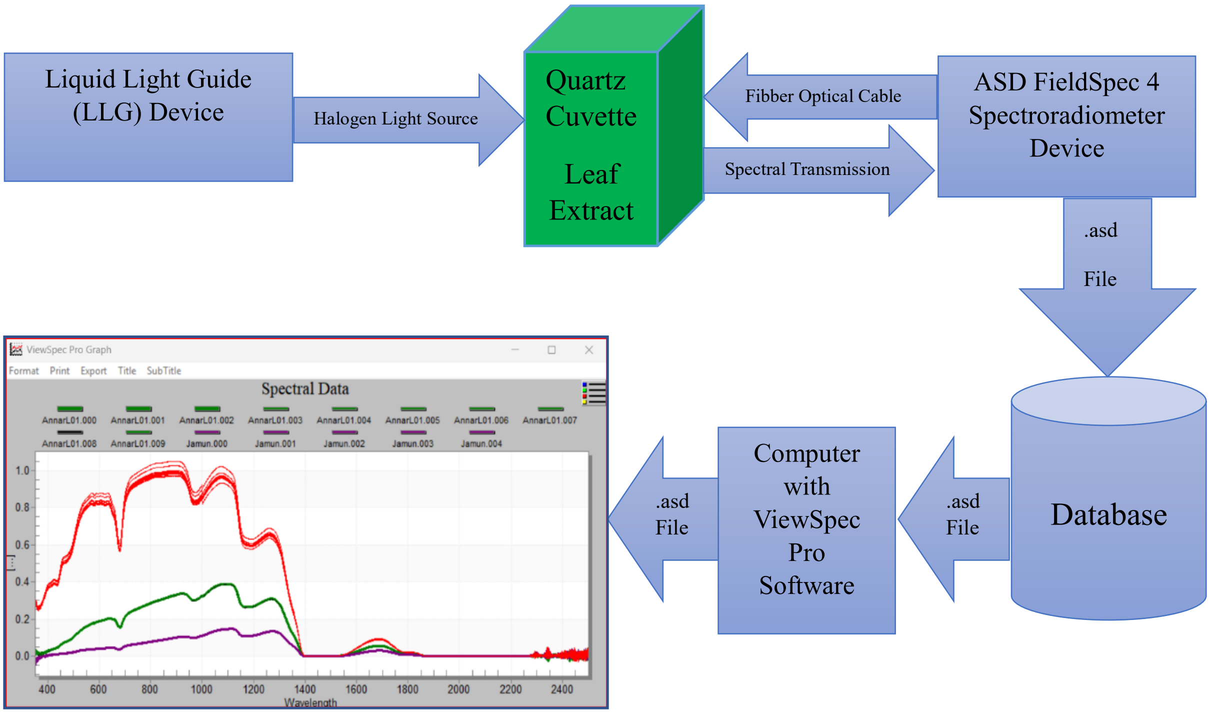

Figure 1: Leaf Liquid Spectral Transmission Data Collection Process

3.1.2 Leaf Sample Collection

Leaves from different plants were pluck under same environmental conditions, but with the fresh and healthy leaves. Biological variability was minimised by using mature leaves that had no physical damage or disease signs. All the leaf samples were washed with distilled water to wash off the dust and surface contaminants and then dried under the lab condition before biochemical extraction.

For each plant species,

- Several genetic samples were collected.
- Every genetic sample was measured individually.
- Ten transmission scans were acquired for every sample.

This strategy improved the repeatability and reliability of the spectral database.

3.1.3 Leaf Extract Preparation

Approximately 2-3 gm of fresh leaves tissue was weighed and crushed using a sterile mortar and pestle. The crushed tissue was mixed with 5 ml of distilled water to extract soluble biochemical compounds. The extract was filtered through laboratory filter paper to remove solid particles, producing a clear solution suitable for transmission spectroscopy. The filtered extract was immediately transferred into a clean quartz cuvette for spectral measurement. The transmitted light emerges from the LLG and is used to illuminate the leaf sample in a quartz cuvette. The selection of a quartz cuvette is due to its optical transparency from UV to SWIR (VIS–NIR) which is not the case of customary glass, where only parts of the UV and infrared region are transparent.

The leaf absorbs, reflects and transmits the incident radiation in three ways:

• Reflection

• Absorption

• Transmission

The interactions are based on the pigments in the leaves, the internal cellular structure, and the water content.

3.1.4 Transmission Spectral Acquisition Using Liquid Light Guide

The Liquid Light Guide was attached between the stabilized halogen light source and the ASD FieldSpec 4 spectroradiometer to obtain the transmission spectra.

Calibration of the instruments was done with the following prior to taking spectra:

• Dark reference calibration

The name might be changed to blank reference (distilled water).

• Instrument optimization

For each piece of biological material,

One of the quartz cuvettes with the leaf extract was inserted in the optical path,

• Ten transmission spectra were recorded.

All the spectra were saved in ASD (.asd) format.

The ASD files were then exported to a CSV file using ViewSpecPro software.

When light shines through the leaf some wavelengths are absorbed and some wavelengths are transmitted.

The information that is carried in the transmitted spectrum includes:

• Chlorophyll concentration
• Carotenoid pigments
• Leaf water content
• Internal leaf structure
• Overall physiological condition

This is the transmitted light that is gathered for a spectral analysis. The transmitted light is gathered by another fibber optic cable, located across from the light source. The fibber optics cable transfers the transmitted radiation to the spectroradiometer while allowing optical losses to be kept to a minimum and ambient light not to contaminate the radiation.Following acquisition, the spectral measurements are saved in the .asd native file format in the ASD instrument.

3.1.5 Spectral Database Distribution

A proposed spectral database was hierarchically organized by plant species, biological samples and repeated scans. The generated **.asd** files are transferred to a database or storage system. The following diagram shows the structure of the database.

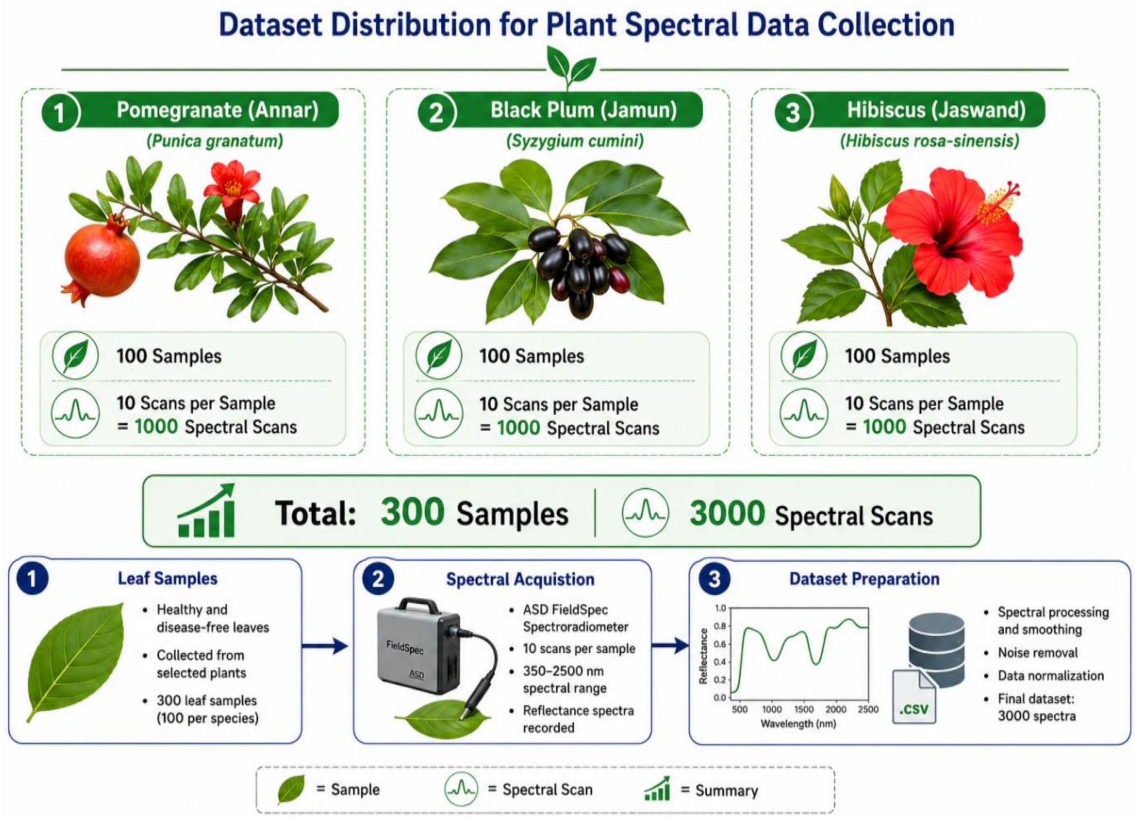

Figure 2 Structure of the database

The average spectrum over 10 repeated scans of each sample was used.

Consequently,

$$S_{avg} = \frac{1}{N}\sum_{i=1}^{N} S_i$$

Where,

- S_i is the ith scan,
- $N = 10$.

The averaged spectrum was considered as the representative spectrum of the biological sample.

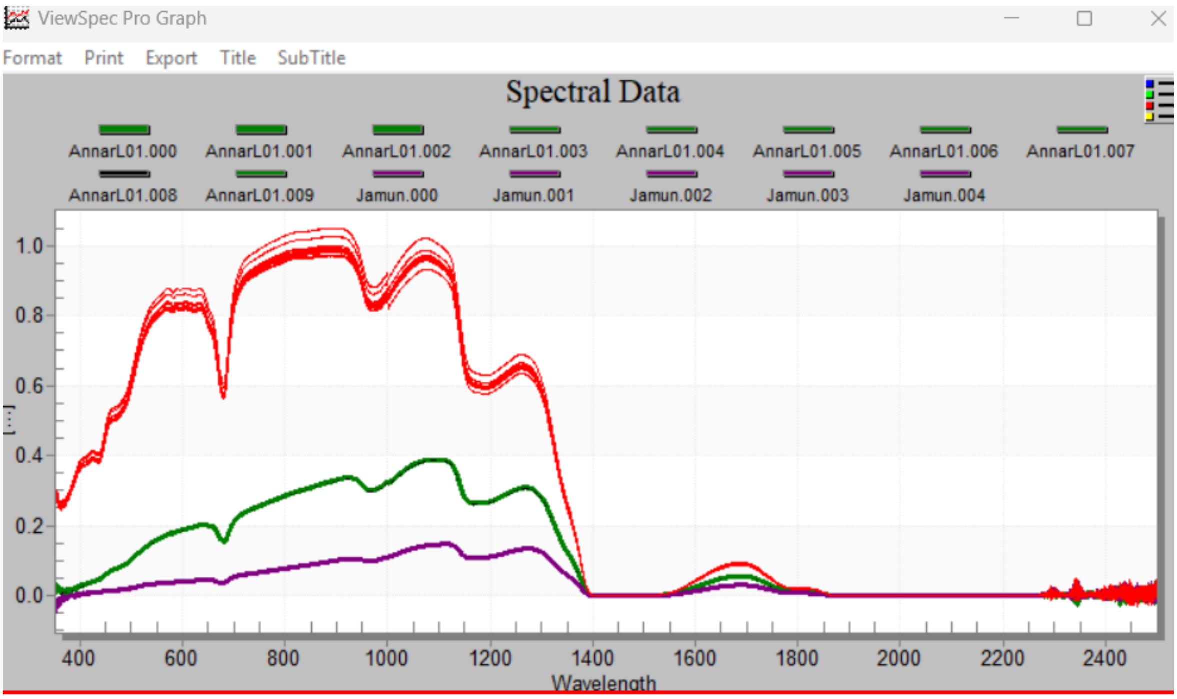

Figure 3: Leaves Spectral Data Before Mean Calculation

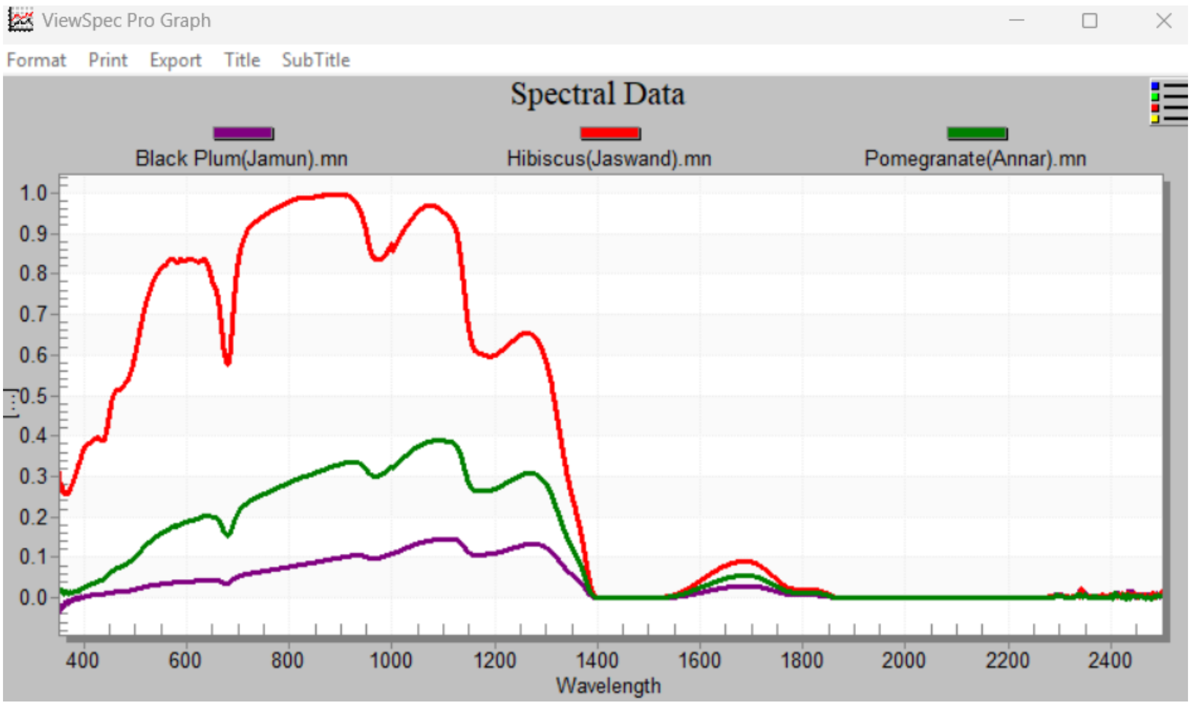

Figure 4: Leaves Spectral Data After Mean Calculation

Spectral data were acquired using an ASD FieldSpec spectroradiometer equipped with a Liquid Light Guide (LLG). For each leaf sample, ten consecutive spectral scans were recorded over the wavelength range of 350–2500 nm under identical experimental conditions. The before mean calculation spectra Figure 3 were examined to evaluate measurement repeatability and spectral consistency. To reduce random instrumental noise and improve signal stability, the 10 scans corresponding to each sample were averaged to generate a single representative spectrum Figure 4. The averaged spectra were subsequently use for preprocessing, feature extraction, and machine learning-based plant species classification.

Conclusion

The development of leaf extract spectral database using Liquid Light Guide (LLG) spectroscopy integrated with an ASD FieldSpec4 spectroradiometer was presented in this paper. A well-defined procedure for leaf sampling, extracting the leaves, obtaining their transmission spectra, compiling them into a database, and pre-processing the spectra was developed to facilitate the creation of a standardized and reproducible spectral database. After taking database we develop spectral library 3000 Scan. The built database is useful for future research in the field of plant spectroscopy and spectral analysis. In future work, the database will be used to examine the extraction of biomarkers from the plant spectra, to develop machine learning and deep learning models to automatically classify the plant species, and to create a chemical fingerprinting model for intelligent plant species identification.

References

1. Wang, M., Lin, H., Lin, H., Du, P., and Zhang, S.: ‘From species to varieties: How modern sequencing technologies are shaping Medicinal Plant Identification’, Genes, 2024, 16, (1), pp. 16
2. Beć, K.B., Grabska, J., Bonn, G.K., Popp, M., and Huck, C.W.: ‘Principles and applications of vibrational spectroscopic imaging in plant science: a review’, Frontiers in Plant Science, 2020, 11, pp. 1226
3. Xia, Y., Lei, H., Zhang, W., Che, T., Yang, Y., Liu, W., Kang, J., Tang, W., and Fan, S.: ‘Recent Advances in Emerging Techniques for Optical Properties Analysis of Fruits and Vegetables: A Review’, Food and Bioprocess Technology, 2026, 19, (6), pp. 264
4. Jiang, Y., Jin, X., Li, G., Ge, H., Yin, Y., Zheng, H., Li, X., and Li, P.: ‘From Fingerprint Spectra to Intelligent Perception: Research Advances in Spectral Techniques for Ginseng Species Identification’, Foods, 2026, 15, (4)
5. J. A. Cope, D. P. Corney, J. Y. Clark, P. Remagnino, and P. Wilkin, "Plant species identification using digital morphometrics: A review," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7562–7573, 2012.
6. S. Wäldchen and P. Mäder, "Plant species identification using computer vision techniques: A systematic literature review," *Archives of Computational Methods in Engineering*, vol. 25, no. 2, pp. 507–543, 2018.
7. P. H. Raven, R. F. Evert, and S. E. Eichhorn, *Biology of Plants*, 8th ed. New York, NY, USA: W. H. Freeman, 2012.
8. H. F. Mahlein, "Plant disease detection by imaging sensors – Parallels and specific demands for precision agriculture and plant phenotyping," *Plant Disease*, vol. 100, no. 2, pp. 241–251, 2016.
9. R. A. Crocombe, "Portable spectroscopy," *Applied Spectroscopy*, vol. 72, no. 12, pp. 1701–1751, 2018.
10. ASD Inc., *FieldSpec® 4 User Manual*, Analytical Spectral Devices Inc., Boulder, CO, USA, 2015.
11. S. Ustin and A. Gamon, "Remote sensing of plant functional traits," *New Phytologist*, vol. 186, no. 4, pp. 795–816, 2010.
12. A. Féret et al., "PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle," *Remote Sensing of Environment*, vol. 193, pp. 204–215, 2017.
13. J. Verrelst, G. Camps-Valls, J. Muñoz-Marí, et al., "Optical remote sensing and machine learning for vegetation studies," *Remote Sensing of Environment*, vol. 216, pp. 60–73, 2018.

14. C. Curran, "Remote sensing of foliar chemistry," *Remote Sensing of Environment*, vol. 30, pp. 271–278, 1989.
15. A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
16. H. Martens and T. Naes, *Multivariate Calibration*. Chichester, U.K.: Wiley, 1989.
17. C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
18. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
19. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
20. R. M. Balabin and S. V. Smirnov, "Variable selection in near-infrared spectroscopy: Benchmarking methods," *Analytica Chimica Acta*, vol. 692, no. 1–2, pp. 63–72, 2011.
21. Ocean Insight, *Liquid Light Guides: Technical Guide*, Ocean Insight Inc., 2022.
22. M. E. Schaepman, S. L. Ustin, A. J. Plaza, T. H. Painter, J. Verrelst, and S. Liang, "Earth system science related imaging spectroscopy—An assessment," *Remote Sensing of Environment*, vol. 113, pp. S123–S137, 2009.