Department of Electrical Engineering, National Tsing Hua University Special Topic on Implementation Research Abstract

Terahertz Time Domain Hyperspectral Imaging for Rapid and Non-destructive Classification among Ginseng Species

太赫茲光譜影像用於人蔘品類分辨

Major Category: Optoelectronics (光電領域)

Group Number: A499

Advisor: 楊尚樺 Shang-Hua

Members: 郭志玲

Research Period: From 2024/9/1 to 2025/3/31.

I. Abstract

This study proposes a novel, streamlined approach that combines terahertz time domain hyperspectral imaging with deep learning to analyze dried ginseng slices, which totally bypasses complex preprocessing and drastically reduces preparation efforts. To mitigate signal variety caused by thickness variation, we introduced a normalization technique to standardize spectral data. By collecting about 12,000 data pixels with sample and employing artificial neuron network for classification, the system achieved accuracy of 97.55%. This method proves to be a rapid, cost-effective and highly tailored approach for real-world scenarios for characterization of traditional Chinese medicine, and can find its place in pharmaceutical and food industries.

II. Introduction

Panax ginseng and Panax quinquefolius, as known as Asian ginseng and American ginseng, are widely utilized in traditional Chinese medicine (TCM) for their diverse therapeutic properties, including but not limited to fatigue alleviation, anti-inflammatory effects, blood circulation regulation, and anti-aging benefits [1]. Despite their taxonomic similarities, there exists distinct therapeutic properties in traditional Chinese medicine (TCM), with AsG classified as "warm" and AmG as "cool", leading to different clinical applications [2]. Furthermore, fluctuations in herb yields and market demand also contribute to regional price variability for AsG and AmG, fostering risks of adulteration in commercial ginseng products [3]. Thus, accurate discrimination between these species is critical for ensuring medicinal efficacy, consumer safety, and combating economic adulteration in commercial products.

Traditional identification methods, such as morphological inspection or sensory evaluation, require extensive expertise and lack reliability for processed forms (e.g., powders, capsules) [4]. Analytical techniques like gas chromatography-mass spectrometry (GC-MS) and high-performance liquid chroma-tography (HPLC), while reliable, involve destructive sample preparation, specialized instrumentation, and skilled operators, limiting their practicality outside laboratory settings [5]. Optical spectroscopy has emerged as a promising alternative due to its non-destructive nature and minimal preprocessing requirements [6]. However, conventional techniques such as Raman and Fourier-transform infrared (FTIR) spectroscopy still face limitations, including sensitivity to sample opacity and the need for homogenized samples or extracts [7]. Even terahertz spectroscopy, which offers unique advantages in molecular fingerprinting and penetration through packaging, typically requires labor-intensive steps like grinding, sieving, and tablet formation to mitigate scattering effects [8, 9]. These preprocessing demands hinder scalability and real-world applicability, particularly in resource-limited environments.

To address these challenges, we used Terahertz Time Domain Hyperspectral Imaging (THz-HSI) combined with deep learning for direct analysis of raw ginseng samples, eliminating the need for destructive or complex preparation. Unlike prior studies that rely on processed tablets or powders, our method leverages raw materials, significantly reducing preprocessing time and complexity. A critical innovation is the introduction of a thickness-agnostic normalization computation to compensate for variations in sample thickness—a major source of measurement inconsistency in optical systems [9]. By generating about 12,000 data pixels with samples and training advanced deep learning models; in fear of overfitting, and to ensure the generality of the results, validation and test dataset on the predictions were made.

III. Results

A. Experimental Setup

THz-HSI measurement was taken place with an Asynchronous Optical Sampling terahertz time-domain spectroscopy (THz-TDS) system, where the THz radiation is focused and transmit through the sample. Following Fig. 1 shows an illustration of the schematic. Difference frequency between the two femto-second lasers was set as 50 Hz, while each waveform was taken with single-shot measurement. Such settings ensure the signals recorded possess dynamic range ~30dB and 1.2 THz bandwidth. Under these settings, a typical ginseng slice in the size of ~ 50mm x 30mm can be scanned with 15 ~ 20 minutes.

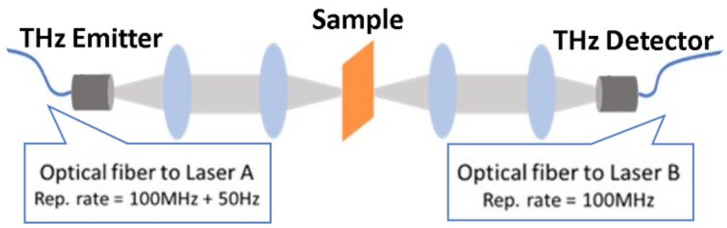

Fig. 1 An illustration of experimental setup of the THz-HIS experiment. The gray blocks indicate the THz optic path under ray optics picture.

B. Computation of Normalization

In our THz-HSI, the transmitted THz optical field is captured as a time-domain waveform $E(t)$. The spectral ampli-tude $\bar{E}(\omega) = |F\{E(t)\}|$. The imaging data were organized into a matrix $\widehat{E_0} = [\overline{E_{\iota J k}}]$, where indices i, j denote horizontal and vertical directions of the THz-HIS, and index k for the values across frequencies, respectively. With the normalization method specified in Eqn. 1, the individual spectrum in each pixel is scaled to its own maximum intensity value. $\hat{E}^N$ is the result of normalization, and $\overline{E_{\iota J}}$ is the whole spectrum in each pixel. This approach emphasizes relative spectral features rather than absolute amplitude variations caused by physical inhomogeneities in the samples, such as thickness fluctuation and porosity, as shown in Fig. 2. This method is also expected to enhance the visibility of intrinsic spectral signatures associated with chemical composition, effectively suppressing confounding factors that could obscure critical discriminative features. [9].

$$\hat{E}^N = \left\{ \frac{\overline{E_{ij}}}{\max\left(\overline{E_{ij}}\right)} \right\} \tag{1}$$

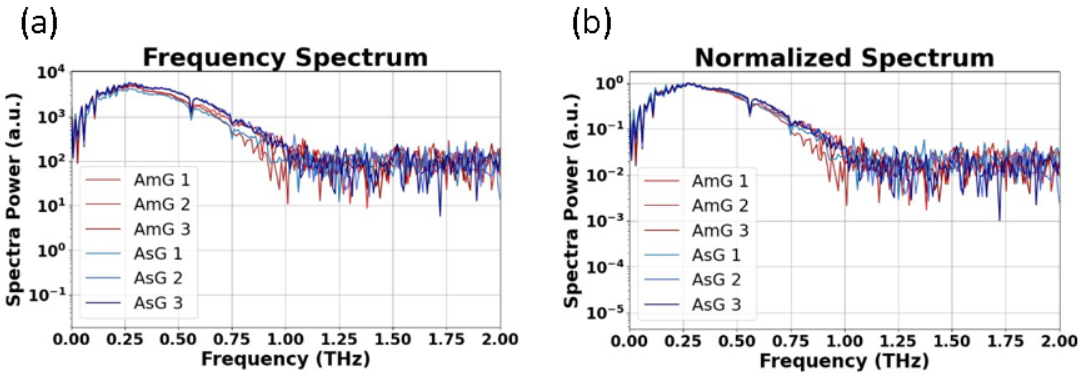

Fig. 2. Comparison and visualization of frequency spectra for multiple datasets (a) before and (b) after normalization.

C. Artificial Neural Network Structure and Analysis

The methodology for utilizing an artificial neural network (ANN) in the identification and prediction of two ginseng species follows a structured architectural design, as illustrated in Fig. 3. The ANN consists of an input layer, six hidden layers with specific node allocations, and an output layer. This architecture processes the normalized dataset through successive hidden layers, ultimately generating a probability matrix with two elements, representing the likelihood of a sample belonging to one of the two ginseng species.

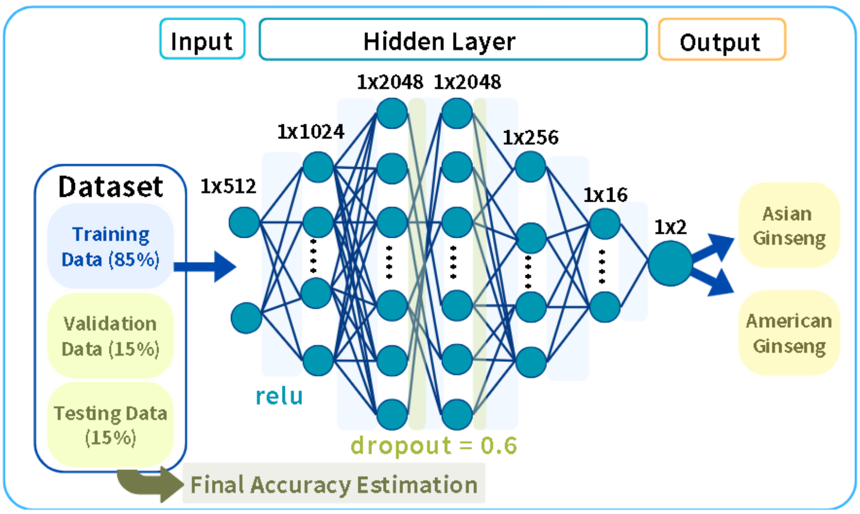

Fig. 3. An illustration of the ANN in this work for classification.

For network training, 70% of the total dataset (8,575 out of 12,250 samples) is used to optimize node values and minimize the loss by comparing predictions to true classifications. This optimization is achieved via backpropagation. To prevent overfitting and enhance generalization, dropout regularization is applied. Additionally, 15% of the dataset (1,837samples) is set aside for validation. Remaining 15% serve as test set.

The model's accuracy is then evaluated on data in test set, to provide a straight-forward evaluation of the performance of this method, as summarized in Table 1. This methodology ensures a balanced and reliable approach to training, validation, and performance evaluation, highlighting the capabilities of ANN for accurate distinguish between the two ginseng species.

Table1. Prediction accuracies of ANN

Prediction Scheme	Loss	Accuracy
0.2 ~ 1.2 THz	0.161	97.552%
0.2 ~ 3.0 THz	0.190	87.595%

Furthermore, for justifying our results, we deliberately include unreliable spectra data from 1.2 ~ 3.0 THz as the comparison group. Lower prediction accuracy is expected in this scheme.

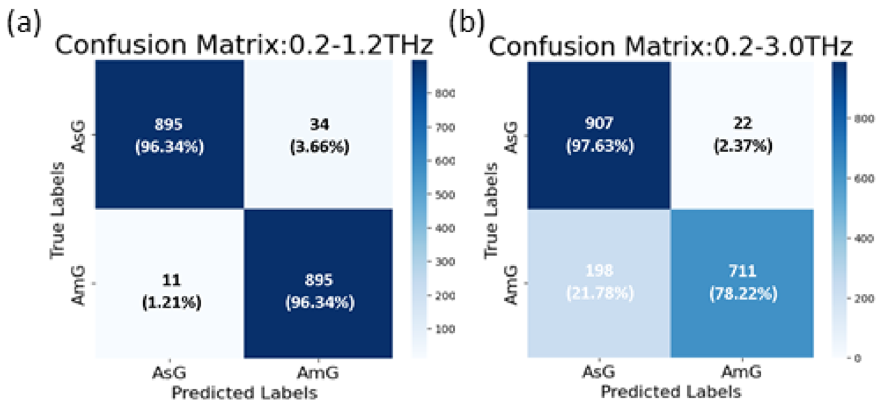

Fig. 4. Confusion matrices comparing prediction accuracies using data from at (a) 0.2 – 1.2 THz and (b) 0.2 – 3.0 THz.

IV. Summary

This study applies THz-SPI coupled with deep learning to distinguish Panax ginseng and Panax quinquefolius directly from dried ginseng slices, bypassing destructive preprocessing. Key innovations include a thickness-agnostic spectral normalization method to suppress sample inhomogeneity. A 6-layer neural network trained on 12,250 spectral datapoints. The model achieved 97.55% accuracy using data within 0.2–1.2 THz range, with confusion matrices confirming robust classification. This approach reduces preprocessing time compared to tablet-based methods, offering a scalable, non-destructive solution for pharmaceutical and food safety industries requiring rapid botanical authentication.

V. References

[1] G. Zhou et al., "Pharmacological effects of ginseng: Multiple constituents and multiple actions on humans," *Am. J. Chin. Med.*, vol. 51, pp. 1085–1104, 2023.

[2] C. F. Chen, W. F. Chiou, and J. T. Zhang, "Comparison of the pharmacological effects of *Panax ginseng* and *Panax quinquefolium*," *Acta Pharmacol. Sin.*, vol. 29, pp. 1103–1108, 2008.

[3] I. H. Baeg and S. H. So, "The world ginseng market and the ginseng (Korea)," *J. Ginseng Res.*, vol. 37, no. 1, pp. 1–7, 2013.

[4] Z. Zhao et al., "Macroscopic identification of Chinese medicinal materials: Traditional experiences and modern understanding," *J. Ethnopharmacol.*, vol. 134, no. 2, pp. 556–564, 2011.

[5] S. H. Baek et al., "Recent methodology in ginseng analysis," *J. Ginseng Res.*, vol. 36, no. 2, pp. 119–134, 2012.

[6] P. J. Potts, "Optical spectrometry: Principles and instrumentation," in *A Handbook of Silicate Rock Analysis*, P. J. Potts, Ed. Dordrecht, Netherlands: Springer, 1987, pp. 77–105.

[7] Y. K. Kwon et al., "Discrimination of cultivation ages and cultivars of ginseng leaves using Fourier transform infrared spectroscopy combined with multivariate analysis," *J. Ginseng Res.*, vol. 38, no. 1, pp. 52–58, 2014.

[8] H. Gu et al., "Identification of *Panax notoginseng* origin using terahertz precision spectroscopy and neural network algorithm," *Talanta*, vol. 274, p. 125968, 2024.

[9] J. C. L. Lin, C. M. Mai, S. S. Wu, W. T. Li, and S. H. Yang, "Deep learning terahertz spectroscopy for non-invasive traditional Chinese medicine identification," in *Proc. Int. Conf. Infrared, Millimeter, Terahertz Waves (IRMMW-THz)*, 2024, pp. 1–2.

[10] Martinez et al., "Chestnut quality classification by THz time-domain hyperspectral imaging combined with unsupervised learning analysis," *Food Control*, vol. 168, p. 110878, 2025.

VI. Review and Reflections

Before I joined the Yang Research Group led by Professor Shang-Hua Yang, I had heard about terahertz (THz) sensing and was drawn to the imaging subgroup due to my curiosity about its potential. Initially, I participated in regular group meetings led by senior members. Their clarity in presentation, efficiency in progress reporting, and rigorous discussion style deeply impressed me. Inspired by their example, I began to read and analyze papers alongside them. Although I was nervous at first, Professor Yang's patient listening and encouragement allowed me the space to gradually improve my presentation and comprehension skills. Over time, I became more confident in interpreting research literature and integrating insights into my own work.

The topic of my project was not pre-assigned—it originated from our team's past experience. A previous senior member had worked on classifying ginseng tablet samples, but the process involved limited data and tedious preprocessing. This motivated me to explore whether we could directly analyze samples via THz scanning, which would streamline the workflow. After discussing various possibilities with senior student Chia-Ming Mai, I settled on a direction that focused on optimizing the data acquisition and classification pipeline. Throughout the project, I experimented with neural network architectures, and learned that the model's performance was highly sensitive to hyperparameters and the number of layers. This required continuous adjustment based on validation results. Paper survey was an essential part of the process—not just passively reading, but actively comparing methods, adapting existing techniques, and identifying gaps I could improve on. This helped me develop a sharper research mindset and stronger critical thinking skills.

Effective project management was key. I maintained short-term goals, validated results step-by-step, and ensured continuous progress through regular discussions with Chia-Ming. Whenever I encountered problems, we would brainstorm solutions together. Time management was also a crucial part of the process. Balancing coursework and research, I gradually developed a workflow that allowed for consistent effort without burnout. In terms of skill acquisition, I actively sought out new tools and methods—from coding neural networks and data visualization, to preparing conference figures and writing reports. Writing documentation was particularly formative. This not only strengthened my writing skills but also helped me grasp how to communicate research clearly and effectively. Moreover, one of the most fulfilling milestones of this journey was becoming the first author of a paper submitted to the International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz). It was an intense but highly rewarding learning experience.

This year-long project greatly improved my ability to manage complex research tasks, think independently, and work collaboratively. I am especially grateful to Professor Yang for fostering an open and encouraging lab environment, and to Chia-Ming for being a generous mentor throughout the journey. Without a doubt, this has been the most transformative and fulfilling experience of my undergraduate years.