

П. Колодочка, А. Ляхнович, М. Ходасевич
(Институт физики НАН Беларуси, г. Минск, Беларусь);

Си Хунджу, PhD, Шу Джей, Сюй Вэй
(Аньхойский Восточно-Китайский НИИ «Фотоэлектроника», г. Уху, КНР)

МНОГОПАРАМЕТРИЧЕСКАЯ КЛАССИФИКАЦИЯ ГЕОГРАФИЧЕСКОГО ПРОИСХОЖДЕНИЯ ЛЕКАРСТВЕННОГО РАСТИТЕЛЬНОГО СЫРЬЯ ПО ТГЦ СПЕКТРАМ ПРОПУСКАНИЯ

Введение. В настоящее время растет интерес к традиционной китайской медицине, успешно использовавшейся на протяжении многих столетий. Основным препятствием для ее широкого применения является отсутствие официальной исследовательской методологии [1].

Набор и состав химических компонентов, входящих в лекарственное растительное сырье (ЛРС), достаточно сложны и могут варьироваться в зависимости от сезона сбора, географического происхождения, окружающего биоразнообразия, процессов заготовки и других факторов. Поэтому для контроля качества ЛРС наиболее приемлемым является метод «отпечатка пальца», который может быть получен с помощью хроматографических, спектроскопических или других методов. «Отпечаток пальца» должен обладать признаками сходства и различия, чтобы однозначно идентифицировать исследуемое лекарственное средство традиционной медицины [2].

С помощью этой характеристики можно определить подлинность ЛРС, даже если концентрации характерных компонентов отличаются для разных образцов.

Целью представленной работы является классификация географического происхождения клубнеобразных корней *Gastrodia*, используемых в традиционной китайской медицине, с помощью анализа многопараметрических данных ТГЦ импульсной спектроскопии поглощения. 50 исследованных образцов клубней лекарственного растения были получены из пяти локаций Китая: Чанбайшань (Changbaishan), Цзиньчжай (Jinzhai), Шэньнунцзя (Shennongjia), Сицзан (Xizang) и Юньнань (Yunnan).

Экспериментальные измерения и вычисление спектров. Аналогично методике, которая описана в [3], ЛРС в виде хорошо высушенных фрагментов клубней измельчалось до однородного порошка, смешивалось с мелкодисперсным политетрафторэтиленом в весовом соотношении 1:20 и просушивалось при температуре 48 °С в течение 4 часов. Размер частиц в смеси не превышал 45 мкм. Прессованием под давлением до 5 т формировались таблетки диаметром 10 мм и толщиной

3,3–4 мм. Измерения временных форм прошедшего через образцы излучения ТГц частотного диапазона выполнялись в Аньхойском Восточно-Китайском НИИ «Фотоэлектроника» (г. Уху, Китай) с использованием спектрометра собственной разработки, схема которого подобна описанной в [4] и приведена на рисунке 1.

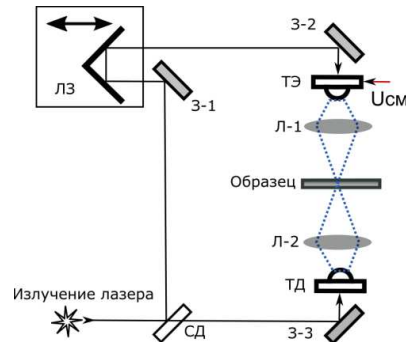


Рисунок 1 – Схема ТГц спектрометра. ЛЗ – линия задержки; ТЭ, ТД – ТГц эмиттер и детектор на основе GaAs; 3-1– 3-3 –зеркала R_{\max} ; СД – делитель; Л-1, Л-2 – линзы

В спектрометре использован фемтосекундный лазер типа FPL-04CFFBW производства компании CALMAR Laser с длиной волны излучения 780 нм. Длительность импульса не превышает 100 фс, частота повторения равна 80 МГц. Фотоантенны стробировались лазерным излучением средней мощностью 16 мВт. Размах импульсного напряжения питания эмиттера U_{cm} составлял 30 В. Ток детектора регистрировался методом синхронного детектирования, при этом для синхронизации использовалась частота модуляции U_{cm} . В спектрометрах временной области при проведении измерений регистрируется временной профиль тока в антенне-детекторе, который пропорционален напряженности электрического поля ТГц излучения. Длительность зарегистрированных профилей составляла 93,33 пс, шаг линии задержки по времени равнялся 0,0267 пс, что соответствует спектральному диапазону 0–18 ТГц при разрешении 10,7 ГГц. Избыточная ширина спектров позволяет более точно определять положение экстремумов и переходов через ноль опорного и сигнального импульсов. Знание этих параметров полезно при выборе вида оконной функции и выполнении преобразования Фурье. Вид характерных временных профилей опорного сигнала и прошедшего через образец *Gastrodia* толщиной 3,74 мм происхождением из Changbaishan, а также спектра пропускания, вычисленного с помощью наложения на временной профиль сигналов оконной функции Ханнинга представлены на рисунке 2. Динамический диапазон спектрометра в максимуме превышает 50 дБ, а частотный диапазон регистрации спектров может составлять не менее 0,1 – 2,0 ТГц для образцов с выбранным составом и геометрией.

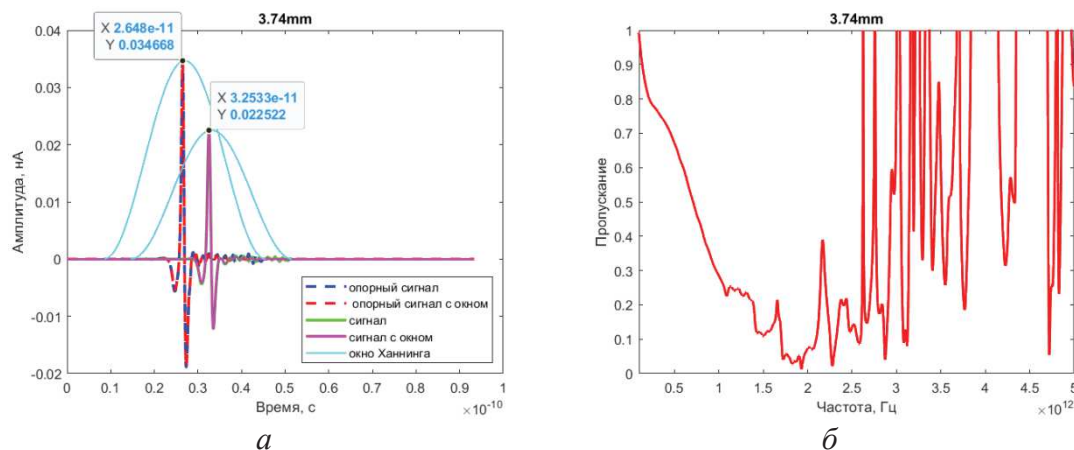


Рисунок 2 – Профили тока в детекторе для опорного сигнала и образца Changbaishan 3,74 мм (а) и спектр пропускания (б)

Многопараметрический кластерный анализ. После вычисления спектров пропускания образцов ЛРС с помощью метода главных компонент [5] определялось маломерное пространство признаков, в котором проводился кластерный анализ. Среди двух методов без учителя (иерархический кластерный анализ и метод k средних) и двух с учителем (деревья классификации и k ближайших соседей kNN [6]) был выбран метод kNN, показавший наибольшую точность классификационной модели. При оптимизации этой модели рассматривалась возможность применения фильтра Савицкого-Голея, выбор положения высокочастотной границы анализируемого спектрального диапазона (0,1–1,5 ТГц), метода предобработки спектров (вычисление первой и второй производных, коррекция базовой линии и несколько видов нормировки), количества главных компонент, количества ближайших соседей, метрики пространства главных компонент и количества спектральных переменных, ранжированных в порядке уменьшения разброса отсчетов в предобработанных спектрах. Основными показателями качества классификационной модели являются точность *Accuracy*, прецизионность *Precision*, чувствительность *Sensitivity* и коэффициент корреляции Мэтьюса *MCC*.

В 6-мерном пространстве главных компонент, построенном по несглаженным фильтром Савицкого-Голея и нормированным на стандартное отклонение 12 спектральным переменным в диапазоне частот 0,1–0,6 ТГц и использующем метрику Махаланобиса, учет двух ближайших соседей в методе kNN позволяет классифицировать 5 рассматриваемых локаций географического происхождения клубней *Gastrodia* со следующими показателями: *Accuracy* = 0,980, *Precision* = 0,980, *Sensitivity* = 0,982 и *MCC* = 0,975.

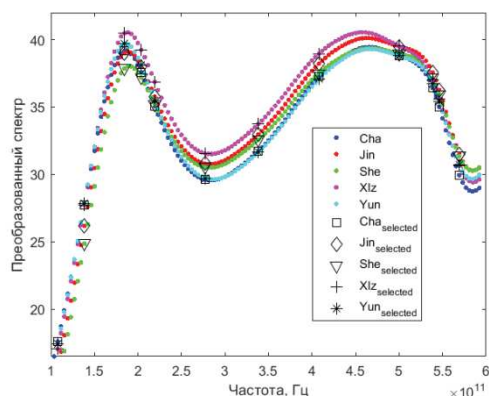


Рисунок 3 – Выбранные частоты для построения оптимального для применения метода kNN пространства главных компонент

Выводы. Продемонстрирована многопараметрическая классификация географического происхождения лекарственного растительного сырья по предобработанным спектрам пропускания частотного диапазона 0,1–0,6 ТГц методом k ближайших соседей. Достигнутая точность 0,98 достаточна для контроля качества средств традиционной китайской медицины.

ЛИТЕРАТУРА

1. Liang, Y-Z. Quality control of herbal medicines / Y-Z. Liang, P. Xie, K. Chan. // *Journal of Chromatography B*. – 2004. – V. 812(1–2). – P. 53–70.
2. Pharmaceutical fingerprinting: Evaluation of neural networks and chemometric techniques for distinguishing among same-product manufacturers / W. J. Welsh [et al.]. // *Anal Chem*. – 1996. – V. 68 (19). – P. 3473–3482.
3. THz-TDS and NIRS Based Detection of Moisture Content for Cattle Feed / Huang J. [et al.] // *Fusion of Front. Phys.* – 2022. – V. 10. – Art. ID 833278. – 13 p.
4. Khodasevich, M. A. Chocolate sample classification by principal component analysis of preprocessed THz transmission spectra / M.A. Khodasevich, A.V. Lyakhnovich, H. Eriklioglu // *Journal of Applied Spectroscopy* – 2022. – V. 89. – N. 2. – P. 251–255.
5. Esbensen, K.H. Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice / K.H. Esbensen, P. Geladi // *Comprehensive Chemometrics* / ed.: S. Brown, R. Tauler, B. Walczak. – 2009. – P. 211–226.
6. Ходасевич, М.А. Многопараметрический подход в методах оптической диагностики: основы и применения / М. А. Ходасевич; Нац. акад. наук Беларуси, Ин-т физики им. Б.И. Степанова. – Минск: Белорусская наука, 2024. – 114 с.