**Применение метода APLPI для классификации чая по запаху**

***Тимакова С.И.1, Кравец К.Ю.1, Азбукин К.В.2***

*Аспирант 2-го года обучения*

*1Институт геохимии и аналитической химии им. В.И. Вернадского Российской академии наук (ГЕОХИ РАН), Москва, Россия*

*2МИРЭА – Российский технологический университет, Москва, Россия*

*E-mail:* *timackovasvet@yandex.ru*

Метод APLPI (Atmospheric Pressure Laser Plasma Ionization), основанный на ионизации органических соединений в условиях воздействия лазерно-индуцированной плазмы при атмосферном давлении, позволяет проводить анализ различных образцов без предварительной подготовки [1]. Такие отличительные особенности, как универсальность ионизации, экспрессность анализа и совместимость с различными типами масс-анализаторов определяют высокий потенциал метода для решения классификационных задач.

В докладе представлены результаты исследования возможности применения метода APLPI для решения задачи классификации чая по запаху. В работе изучены приобретенные коммерчески образцы листового чая, относящиеся к трем группам: черный, зеленый и улун. Каждая группа представлена тремя товарным наименованиям. Разработанный ионный источник APLPI использовали в сочетании с масс-анализатором типа Q-TOF (MX5311, ИАП РАН). Генерация плазмы осуществлялась путем воздействия сфокусированного импульсного излучения Nd:YAG лазера на поверхность металлической мишени, находящейся вблизи заборника пробы масс-спектрометра. Мишень помещена в камеру, находящуюся при атмосферном давлении. Измерения проходили в атмосфере аргона. Образцы чая помещали в виалу, нагреваемую примерно до 55⁰С и расположенную в газовом тракте ионного источника. Каждый образец проанализирован 21 раз. Результирующие масс-спектры преобразованы в массив данных и случайным образом разделены на стратифицированные по классам тестовый и обучающий массив (в соотношении 1:2, соответственно). Для обработки данных использовался статистический пакет «Orange Data mining», а также пакеты «scikit-learn» и «Pandas». В качестве методов кластеризации без учителя использовались метод главных компонент, метод независимых компонент, K-средних. В качестве методов машинного обучения использованы метод опорных векторов (SVM), градиентный бустинг, в реализации Catboost, метод случайного леса (Random Forest), наивная Баесовская классификация (Naive Bayes classifier).

Показано, что методы кластеризации без учителя неэффективны для разделения на основе полученных данных вследствие их высокой внутренней вариативности и большого количества неинформативных признаков. Для решения задачи классификации чая необходимо использовать методы машинного обучения. Наивысшая точность определения класса для полученного массива данных достигнута при использовании метода опорных векторов, который обеспечивает точность классификации (Classification Accuracy) 0.94 (для разделения чая по товарным наименованиям) и 0.98 (для разделения по степени ферментации). Изучена возможность применения методов машинного обучения для отбора значимых признаков. Наиболее эффективным методом для решения этой задачи является метод случайного леса. Отбор признаков этим методом позволил довести эффективность классификации по товарным наименованиям до 1.

**Литература**

1. Кравец К.Ю., Симановский Я.О., Гречников А.А. Экспресс-анализ лекарственных средств методом масс-спектрометрии с ионизацией излучением лазерной плазмы при атмосферном давлении // Биомедицинская радиоэлектроника. 2020. Т. 23. №. 2. С. 15–21.