**Использование кластеризации данных в методе ГХ-МС при анализе произведений живописи**

***Малышев А.Н., Данилов Д.А., Хорькова А.Н., Ильичев Д.В.***

*Студент, 1 курса магистратуры*

*УрФУ им. Б.Н. Ельцина, Екатеринбург, Россия*

*E-mail: Malyshev.k1b@gmail.com*

Кластеризация – это многомерная статистическая процедура, основана на объединении множества объектов на подмножества (кластеры) по заданному критерию. Данная процедура получила широкое распространение и успешно используется в обработке физико-химических и биологических данных [1,2].

Целью исследования являлась проверка возможности использования кластеризации данных в методе ГХ-МС при анализе образцов картин, которые представляли собой верхний слой.

Объектом исследования является верхний слой картин. В большей степени он представляет собой покрывной лак и масла. Также в верхний слой могут входить частицы органических и неорганических красителей.

По результату сравнения состава некоторых лаков на природной основе, можно заметить, что у каждого лака есть свои собственные характерные соединения. Опираясь на подобные соединения, в дальнейшем происходил анализ реальных образцов и идентификация используемого лака написанной картины. В свою очередь, например, по соотношению пальметиновой (С16:0) и стеариновой (С18:0) кислот возможно определить природу лака[3].

В связи с чем для достижения цели исследования проводили анализ продуктов экстракции и жирных кислот. Соответственно, аналитами являлись жирные кислоты и многочисленные органические соединения получаемые по результату экстракции.

Анализ проводился на хроматографе Perkin Elmer Clarus 600 с масс-спектрометром Perkin Elmer Clarus 600T, капиллярная колонка Elite-5MS (30 м, 0,25 мм, 0,25 мкм).

Кластеризацию полученных данных проводили, используя язык R (версия программы R 4.2.2) [4]. Базу данных для обработки формировали, используя ПО Microsoft Excel.

В случае продуктов экстракции выполняли иерархическую агломеративную кластеризацию. Были проверены 2 варианта: 1) учет всех индивидуальных соединений, где наличие соединения в лаке обозначалось единицей, а их отсутствие - нуль; 2) все соединения, которые были обнаружены, объединялись по соответствующим им органическим классам. В результате обработки данных пришли к выводу о том, что наиболее достоверным является второй вариант.

К результатам жирокислотного анализа также применили иерархическую агломеративную кластеризацию. В данном случае наилучший вариант - учитывать именно соотношения различных жирных кислот.

В дальнейшем планируется провести кластеризацию продуктов экстракции картин с помощью функции тепловой карты (Heatmap) в R [5].

**Литература**

1. Böcker A. et al. A Hierarchical Clustering Approach for Large Compound Libraries // J. Chem. Inf. Model. 2005. Vol. 45, № 4. P. 807–815.

2. Mocz G. Fuzzy cluster analysis of simple physicochemical properties of amino acids for recognizing secondary structure in proteins: Fuzzy clustering for structure prediction // Protein Science. 1995. Vol. 4, № 6. P. 1178–1187.

3. Косолапов А.И. Естественнонаучные методы в экспертизе произведений искусства. // СПб: Изд-во Гос. Эрмитажа. 2010. 170 с.

4. Guha R. Chemical Informatics Functionality in R // J. Stat. Soft. 2007. Vol. 18, № 5. P. 1-16.

5. Kassambara A. Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning (Multivariate Analysis) // STHDA. 2017. P. 187.