

Д. Сатыбалдина<sup>✉</sup>, А.Исайнова, Н.Глазырина, А. Шагабудинов

Евразийский национальный университет имени им. Л.Н.Гумилева, Астана, Казахстан  
E-mail: satybaldina\_dzh@enu.kz

## СТАТИСТИЧЕСКИЕ МЕТОДЫ ИЗВЛЕЧЕНИЯ ПРИЗНАКОВ ДЛЯ ОБУЧЕНИЯ МУЛЬТИСЕНСОРНОЙ ГАЗОАНАЛИТИЧЕСКОЙ СИСТЕМЫ

**Аннотация.** Мультисенсорные газоаналитические системы типа «электронный нос» используются для детектирования утечки вредных газов, обнаружения опасных веществ, контроля качества пищевых продуктов, в медицинской диагностике. В таких системах используются алгоритмы машинного обучения для реализации распознавания газов на основе близости сенсорных откликов на исследуемые аналиты и информационных признаков, использованных на этапе обучения. Для повышения точности классификации запахов необходимо развивать методы анализа сенсорных откликов для расширения наборов информационных признаков. В работе представлены результаты статистической обработки экспериментальных данных по изменению электрических характеристик восьми датчиков «электронного носа» как реакций на присутствие в воздухе летучих органических соединений от образцов сырой нефти. Показано, что анализ описательных статистических характеристик, методы регрессионного и кластерного анализа позволяют провести отбор коррелирующих факторов, ранжировать факторы по степени их влияния на сенсорные отклики.

**Ключевые слова.** Газовый сенсор, извлечение признаков, искусственное обоняние, классификация запахов, корреляционный анализ, машинное обучение; описательные статистики, регрессионный анализ.

### Введение.

Сенсорная технология искусственного обоняния началась с изобретения первой газовой мультисенсорной матрицы в 1982 году [1]. Газоаналитическая система типа «электронный нос» обычно состоит из мультисенсорной матрицы, блока предобработки информации, блока обработки сенсорных данных и библиотеки сенсорных откликов на известные газообразные вещества [2]. Матрица перекрёстно-реагирующих сенсоров может быть представлена линейкой постепенно различающихся сенсоров, выбранных для реагирования на широкий спектр химических классов и распознавания аналитов и их смесей. Выходные данные от отдельных сенсоров собираются вместе и интегрируются для создания уникального цифрового отклика, наподобие отпечатка пальца, что позволяет осуществить идентификацию и классификацию аналитов. Достижения в развитии технологии газовых сенсоров, электроники, биохимии и искусственного интеллекта позволили разработать устройства, способные измерять и характеризовать летучие ароматы, выделяемые из множества источников, для многочисленных применений [3, 4].

Одной из важных проблем сенсорной технологии искусственного обоняния является необходимость проведения множества экспериментов для целевых газов, проведения процедур извлечения признаков, сбора датасетов для реализации этапов обучения и тестирования классификатора газа на основе алгоритмов машинного обучения или искусственной нейронной сети.

Ранее нами были представлены результаты исследований по созданию газоаналитической системы типа «электронный нос» с использованием восьми металл-оксидных сенсоров и блоками обработки сенсорных данных на Python, получены оценки

точности распознавания модельных аналитов при варьировании исследуемых газов и их концентраций [5]. После серии экспериментов с газообразными веществами собраны базы сенсорных данных. Возможность прямого использования этой базы данных на другой необученной газоаналитической системе осложняется трудностью создания двух идентичных сенсорных систем, так как трудно получить одинаковые газовые сенсоры с точно воспроизводимыми параметрами [2]. В связи с этим нами была модель переноса базы сенсорных данных, которая позволяет сократить время на обучение каждой новой газоаналитической системы, используя дополнительные процедуры «тонкой» настройки системы классификации для минимизации эффектов «неидентичности» газовых сенсоров [6].

В настоящей работе представлены результаты исследований по развитию методов преобразования сырых сенсорных данных в наборы информативных признаков для повышения точности распознавания газов, выбора информативных факторов, определения степени важности откликов отдельных сенсоров «электронного носа». С этой целью проведена статистическая обработка экспериментальных результатов с использованием пакета Statistica (64-битная версия v. 12.5.192.7, TIBCO, USA) [7]. Применимость статистических параметров для извлечения информативных признаков и последующего использования алгоритмами машинного обучения продемонстрировано в ряде работ зарубежных исследований [8, 9]. Полученные нами результаты также указывают на то, что процедуры статистической обработки сенсорных данных позволяют улучшить качество наборов информационных признаков для обучения мультисенсорных газоаналитических систем.

#### **Материалы и методы.**

В таблице 1 представлены данные по исследованным образцам сырой нефти с девяти месторождений Казахстана. Несмотря на то, что все месторождения находятся в одном регионе, на западе страны, образцы сырой нефти различаются по своим физико-химическим характеристикам. Выбор сырой нефти в качестве образцов исследования связан с проблемой нефтяных загрязнений почв и водных бассейнов Казахстана. Нефть содержит углеводороды, которые имеют низкую температуру кипения и классифицируются как летучие органические соединения (ЛОС) с известным опасным воздействием на здоровье человека и воздушную экосистему [4]. Газоаналитическая система типа «электронный нос» может детектировать ЛОС и быть полезной при обнаружении и мониторинге разливов нефти при добыче, транспортировке и хранении.

В работе использована собранная ранее газоаналитическая система типа «электронный нос» [5]. На рисунке 1 представлена схема экспериментальной установки по исследованию изменения электрических характеристик сенсоров газоаналитической системы на присутствие ЛОС от образцов сырой нефти. Характеристики и обозначение сенсоров «электронного носа» показаны в таблице 2.

Схема на рисунке 1 включает переключатель потока газа (ПП, регуляторы расхода газа (РРГ), контейнер с образцом. К газоаналитической системе подводится питание, блок управления потоками; газоаналитическая система подключается к компьютеру, в котором установлено разработанное для управления и сбора и анализа данных программное обеспечение.

ПП управляется программой и позволяет направлять потоки сухого воздуха (линия 1) и воздуха с анализируемыми парами (линия 2 через контейнер с образцом) поочередно на сенсоры «электронного носа» с выбранным временным интервалом (зависит от времен отклика и восстановления сенсоров). Сенсорные данные собираются для анализа с помощью программ, установленных на персональном компьютере.

Протокол эксперимента включает измерение электрических характеристик сенсора на сухом воздухе в течение 30 минут, затем на воздухе, проходящем через образец с

сырой нефтью (500 мкл) в течение 15 минут, затем переключение потока на синтетический воздух на 30 минут для стабилизации сенсоров, затем все повторяется два раза измерения сенсорных откликов в присутствии ЛОК от нефти, между ними 30 сухого воздуха.

Таблица 1 - Описание образцов сырой нефти

Обозначение образца	Наименование месторождения, регион Казахстана	Оператор месторождения	Краткие сведения о физико-химических свойствах образца сырой нефти [10]
oil_n1	Алибекмола, Актюбинская область	ТОО «Казахойл Актобе»	Плотность нефти - 0,835 г/см <sup>3</sup> , содержит сернистые соединения (1,14-1,81%) и парафины (3,2-8,18%).
oil_n2	Алибек Южный, Актюбинская область	АО «Каспий Нефть ТМЕ»	Плотность нефти - 0,842 г/см <sup>3</sup> , содержит сернистые соединения (1,33%) и парафины (6,55%).
oil_n3	Восточный Жагабулак, Актюбинская область	ТОО «Арал Петролеум	Легкая нефть (плотность 0,838 г/см <sup>3</sup> ) с высоким содержанием сернистых соединений (2.3%).
oil_n4	Кардасын северный, Актюбинская область	ТОО «Сагиз Петролеум Компани»	Тяжелая нефть (плотность 0,945 г/см <sup>3</sup> ), имеет в составе парафины (1,57%), сера (0,2%), смолы (15,6%), асфальтены (1,55%).
oil_n5	Сарыкумак, в границах Атырауской и Актюбинской областей	ТОО «Сагиз Петролеум Компани»	Тяжелая нефть (плотность 0,896 г/см <sup>3</sup> ), имеет в составе парафины (2,7%), сера (0,4%), смолы (14,4%), асфальтены (2,4%).
oil_n6	Жанажол, Актюбинская область	АО «СНПС - Актобемунайгаз»	Легкая нефть (плотность 0,809-0,827 г/см <sup>3</sup> ), маловязкая, сернистая (0,7-1,11%), парафинистая (4,9-7,1%). Содержание силикагелевых смол 4,23-6,8%, асфальтенов 0,43-1,78%.
oil_n7	Жанаталап, Атырауская область	АО «Эмба мунайгаз»	Плотность нефти 0,814-0,892 г/см <sup>3</sup> . Относится к малосернистой (0,17-0,38%) и малопарафинистой (0,31-1,01%) нефти.
oil_n8	Чинаревское, Западно-Казахстанская область	ТОО «Жаик мунай»	Легкая нефть (плотность 0,8164 г/см <sup>3</sup> ), малосернистая (0,32 %), малосмолистая (2,7 %) и парафинистая (5,6 %).
oil_n9	Сазанкурак, Атырауская область	ТОО «Сазанкурак»	Тяжелая нефть (плотность 0,9245 г/см <sup>3</sup> ), имеет в составе парафины (5,08%), сернистые соединения (1,2%), силикагелевые смолы (до 41,02%), асфальтены (1,73%).

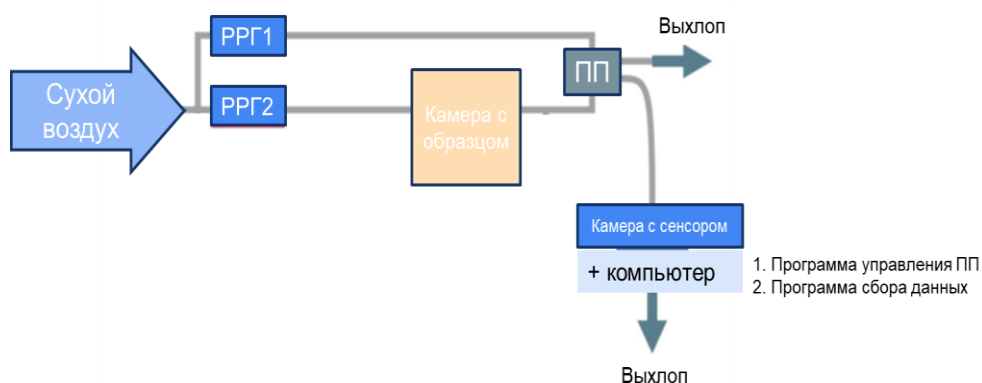


Рисунок 1 - Установка с регулируемой подачей газов, объединенная с прибором типа «электронный нос»

Таблица 2 - Обозначение сенсоров

Наименование марки сенсора	Целевые газы	Обозначение датчика
MQ-7	CO	R1
MQ-3	Alcohol	R2
MQ-5	LPG	R3
MQ-2	Smoke	R4
MQ-9	CO, CH <sub>4</sub> , LPG	R5
MQ-8	H <sub>2</sub>	R6
MQ-135	NH <sub>3</sub> , CO <sub>2</sub> , NO <sub>2</sub>	R7
MQ-4	CH <sub>4</sub>	R8

### Результаты.

На рисунке 2 представлен пример временной зависимости сенсорных откликов, полученной по протоколу эксперимента. Предполагая наличие взаимосвязи между сенсорными реакциями «электронного носа» с различиями в физико-химических свойствах экспериментальных образцов, чувствительности разных сенсоров к ЛОС от сырой нефти, мы провели статистический анализ сенсорных данных: описательные статистики (среднее, медиана, мода, дисперсия, гистограммы распределения), множественный корреляционный анализ и построение многомерной регрессионной модели).

Сенсорные отклики для 9 образцов сырой нефти и воздуха представлены в виде помеченных данных (по 20 векторов на каждый образец, значения 8 газовых сенсоров, соответствующих фазе насыщения и значения температуры и влажности, последний столбец – обозначение образца нефти).

Для проведения регрессионного анализа зависимости сенсорных откликов от состава сырой нефти используются данные хроматографии - масс-спектрометрии образцов сырой нефти, предоставленные зарубежным партнером проекта.

На рисунках 3 (а, б) приведены примеры гистограмм нормального распределения сенсорных откликов для двух образцов нефти, различающихся по плотности и содержанию парафинов. Тесты на нормальность распределения проведены по двум критериям. На гистограммах распределения указаны уровни значимости  $p$  по критерию Колмогорова-Смирнова (the Kolmogorov-Smirnov test, K-S) и по критерию Шапиро-Уилка (the Shapiro-Wilk test). Отклонение от нормального распределения считается существенным при значении  $p < 0,05$ .

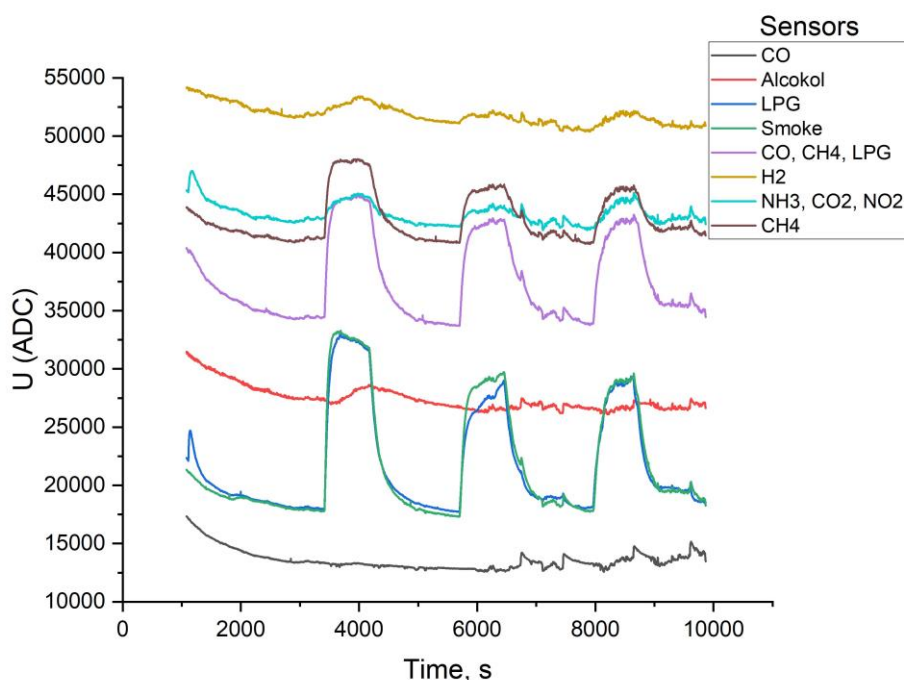
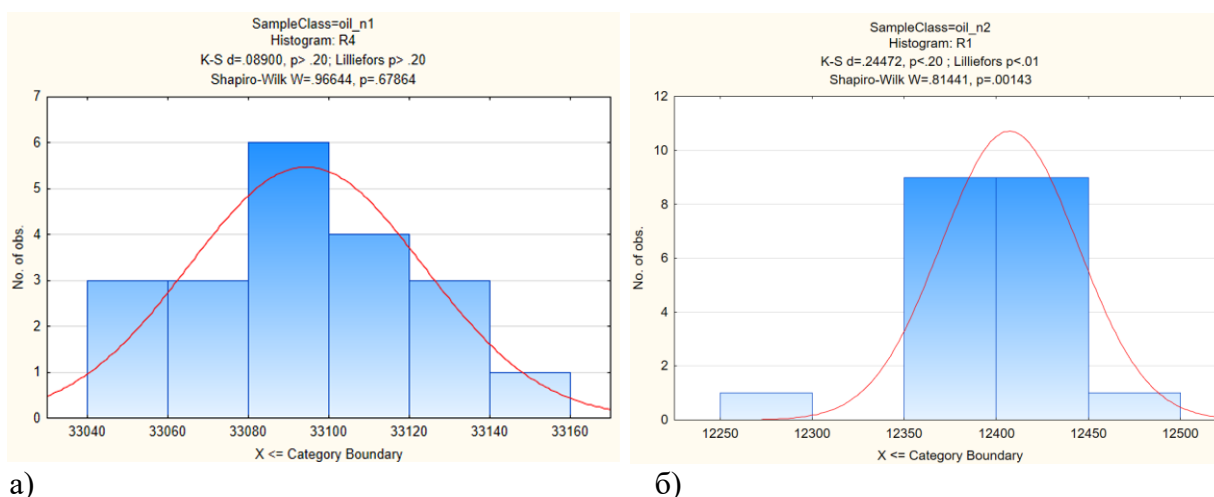


Рисунок 2 - Результаты трех измерений сенсорных откликов в присутствии ЛОК от образца нефти #1 (месторождение Алибекмола)

Для всех сенсорных откликов и всех образцов нефти (то есть для всех функций распределения значений сенсорных откликов,  $8 \times 9 = 72$  выборки по 20 значений) уровень значимости по критерию Колмогорова-Смирнова  $p > 0,05$ . Только для 4-х выборок сенсоров из 72 критерий Шапиро-Уилка имеет значимую вероятность ошибки (см. пример на рисунке 3б).



а)

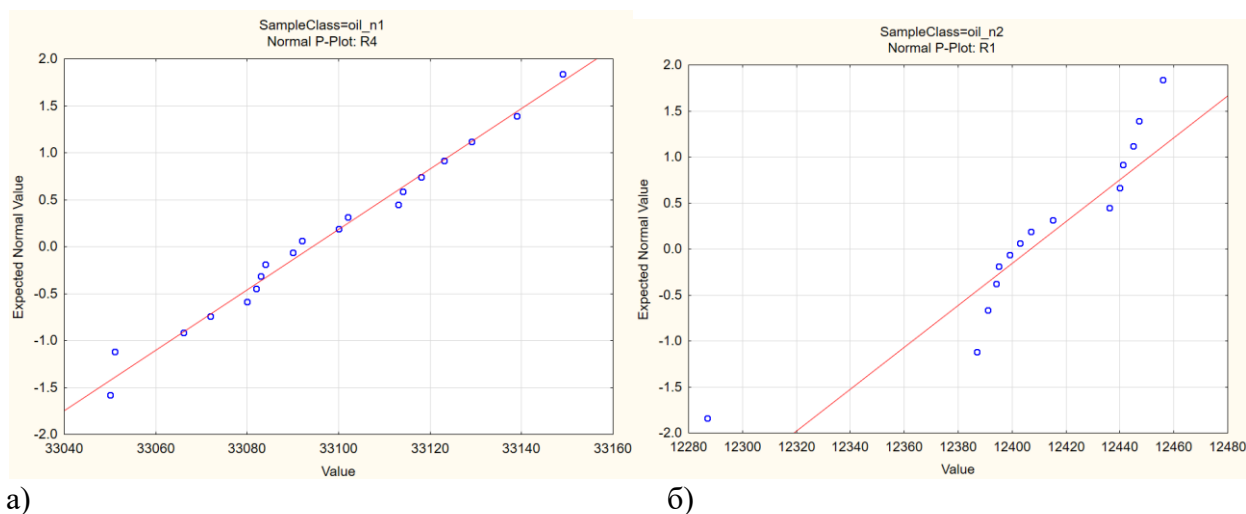
б)

а) для сенсора R4 и образца нефти oil\_n1;  
б) для сенсора R1 и образца нефти oil\_n2.

Рисунок 3 - Гистограммы нормального распределения

На нормально-вероятностных графиках (см. примеры на рисунке 4а) экспериментальные точки лежат вдоль теоретической прямой, что также указывает, что значения переменных достаточно хорошо подчиняются нормальному распределению и можно применять параметрические тесты. Исключение составляют случаи при

отклонении от нормального распределения по критерию Шапиро-Уилка (рисунок 4 б). Аналогичные результаты получены для другого графического теста на нормальность – диаграмм размаха.



а) для сенсора R4 и образца нефти oil\_n1;  
б) для сенсора R1 и образца нефти oil\_n2.  
Рисунок 4. Нормально-вероятностные графики

В целом, можно констатировать, что выборки сенсорных откликов подчиняются нормальному распределению значений. Следовательно, для выявления корреляции между исследуемыми переменными можно применить коэффициент корреляции Пирсона.

### Обсуждение.

Согласно данным термической газовой хроматографии-масс-спектрометрии образцы сырой нефти содержит разветвленные и нормальные алканы, циклоалканы и бензолы (ароматические соединения, арены). Получены результаты вычисления парной корреляции между всеми сенсорными откликами и групповым составом нефти (таблица 3). Курсивом выделены корреляции, которые являются значительными (при  $p < 0,05$ ).

Таблица 3 - Корреляционная матрица для всех факторов и откликов

Variable	Correlations (oil_for_1_9.sta)							
	R1	R2	R3	R4	R5	R6	R7	R8
Alkane	<i>0.391598</i>	<i>-0.272108</i>	<i>-0.800496</i>	<i>-0.773893</i>	<i>-0.824868</i>	<i>-0.801175</i>	<i>-0.641411</i>	<i>-0.747633</i>
Aroma	<i>-0.525409</i>	<i>0.443612</i>	<i>0.923034</i>	<i>0.873743</i>	<i>0.819834</i>	<i>0.895672</i>	<i>0.867346</i>	<i>0.846958</i>
Cyclopentane	<i>-0.025873</i>	<i>0.193728</i>	<i>0.331077</i>	<i>0.305951</i>	<i>0.414240</i>	<i>0.368110</i>	<i>0.294417</i>	<i>0.290279</i>
Cyclohexane	<i>-0.142467</i>	<i>-0.075949</i>	<i>0.246767</i>	<i>0.276131</i>	<i>0.406429</i>	<i>0.251699</i>	<i>-0.012157</i>	<i>0.260462</i>
Other	<i>0.115257</i>	<i>-0.466454</i>	<i>0.506538</i>	<i>0.470404</i>	<i>0.409589</i>	<i>0.607415</i>	<i>0.514169</i>	<i>0.506905</i>

В таблице 4 представлены пример построения многомерной регрессионной модели для сенсора R1. При выборе параметров регрессии были включены все независимые переменные, но итоговая регрессионная модель содержит не все. Полученное значение  $p$ -уровня  $< 0,05$ , что указывает на значимость переменных в регрессионной модели. Значения коэффициентов корреляции указывают на наличие средней связи.

Таблица 4 - Regression Summary for Dependent Variable: R1

N=180	R= .79784709 R <sup>2</sup> = .63655998 Adjusted R <sup>2</sup> = .62825278 F(4,175)=76.627 p<0.0000 Std.error of estimate: 423.9					
	b* (Beta)	Std.Err. of b*	b	Std.Err. of b	t(6)	p-value
Intercept			21955.58	1000.170	21.9519	0.000000
Aromas	-2.58241	0.224648	-125.48	10.915	-11.4954	0.000000
Other	0.36118	0.058696	217.52	35.349	6.1535	0.000000
Cyclohexanes	-2.21147	0.219143	-152.23	15.085	-10.0914	0.000000
Alkanes	-2.75015	0.320323	-83.84	9.765	-8.5855	0.000000

Для сенсора R2 получены более высокие значения коэффициента корреляции R, коэффициента детерминации R<sup>2</sup> и скорректированного коэффициента Adjusted R<sup>2</sup>, которые указывают на наличие сильной связи (таблица 5).

Таблица 5 - Regression Summary for Dependent Variable: R2

N=180	R= .89916394 R <sup>2</sup> = .80849580 Adjusted R <sup>2</sup> = .80411856 F(4,175)=184.70 p<0.0000 Std.Error of estimate: 1502.4					
	b* (Beta)	Std.Err. of b*	b	Std.Err. of b	t(6)	p-value
Intercept			65086.40	3544.629	18.3620	0.000000
Other	-0.99700	0.042607	-2931.50	125.277	-23.4001	0.000000
Alkanes	-2.32564	0.232520	-346.15	34.609	-10.0019	0.000000
Cyclohexanes	-1.22672	0.159075	-412.27	53.460	-7.7116	0.000000
Aromas	-0.86236	0.163070	-204.57	38.684	-5.2883	0.000000

Данные парной корреляции и регрессионного анализа позволили определить влияние каждого фактора (химического компонента) на отклики сенсоров, а также чувствительность отдельных сенсоров к ЛОС от химических составляющих нефти (таблица 5).

Таблица 5 - Влияние состава сырой нефти в образцах факторов на сенсорные отклики

Variable	Фактор входит или нет в модель регрессии							
	R1	R2	R3	R4	R5	R6	R7	R8
Alkane	+	+	-	-	+	+	-	-
Aroma	+	+	+	+	-	+	+	+
Cyclopentane	-	-	+	+	+	-	+	+
Cyclohexane	+	+	+	+	+	+	+	+
Other	+	+	+	+	+	+	+	+

### Заклучение.

В работе представлены результаты научных исследований по статистическому анализу экспериментальных данных по измерению сенсорных откликов мультисенсорной газоаналитической системы типа «электронный нос» в присутствии летучих органических соединений от образцов сырой нефти с девяти месторождений западного Казахстана.

Проведенные экспериментальные исследования показали, во-первых, возможности использования систем искусственного обоняния для различения разных образцов нефти по запахам. Во-вторых, выявлено, что определение степени важности откликов отдельных сенсоров «электронного носа», выбор влияющих факторов на основе метода анализа описательных статистик, множественного корреляционного анализа и многомерного регрессионного моделирования позволяет скорректировать процедуры выделения

информационных признаков для задач обучения системы распознавания газов, улучшить наборы информативных признаков для повышения точности классификации исследуемых образцов.

Будущие исследования связаны с проведением дополнительных экспериментов по расширению базы газовых паттернов, апробированием модели переноса базы знаний на необученные мультисенсорные газоаналитические системы типа «электронный нос» в целях сокращения этапа обучения и улучшения условий для их серийного производства.

**Благодарности.** Данная работа выполнена при финансовой поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан (грант AP14872171). Коллектив авторов благодарит Абишева Е. (АО «Каспий Нефть ТМЕ») за предоставленные образцы сырой нефти с месторождений Казахстана; Федорова Ф.С. (Сколковский институт науки и технологий, РФ) за предоставленные данные хроматографии - масс-спектрометрии образцов сырой нефти

## ЛИТЕРАТУРА

[1] Persaud, K.C.; Dodd, G. Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose // *Nature*. – 1982. – V. 299. – Pp. 352-355.

[2] Karakaya D., Ulucan O., Turkan M. Electronic nose and its applications: A survey // *International journal of Automation and Computing*. – 2020. – V. 17. – №. 2. – Pp. 179-209.

[3] Zaytsev V., Fedorov, F. S., Satybaldina D. et al. Rapid and accurate quality assessment method of recycled food plastics VOCs by electronic nose based on Al-doped zinc oxide // *Journal of Cleaner Production*. – 2023. – V. 418. – Paper N. 138042. <https://doi.org/10.1016/j.jclepro.2023.138042>

[4] Satybaldina, D., Baydeldinov, M., Issainova, A., Alseitov O., Konyrkhanova A., Akhmetova, Z., Seilov, S. Artificial Olfactory System for Distinguishing Oil-Contaminated Soils // *WSEAS Transactions on Environment and Development*. – 2023. V. 19. – Pp. 951–960.

[5] Сатыбалдина, Д., Исайнова, А., Альсеитов, О., Байдельдинов, М., Сеилов, Ш. Создание газоаналитической мультисенсорной системы типа «электронный нос» и исследование её производительности для задач экологического мониторинга // *Вестник КазАТК*. – 2023. – № 125(2). – С. 315–323.

[6] Сатыбалдина Д., Калымова К., Сыдыков Д. Применение трансфера обучения нейронных сетей при классификации изображений // *Вестник КазНПУ им. Абая, серия «Физико-математические науки»*. -2023. - №1 (81). - С.159-169.

[7] TIBCO Data Science – Statistica. Available online: <https://www.tibco.com/products/data-science/> (дата обращения 05 декабря 2023 года).

[8] Suhaidi M., Kadir R. A., Tiun S. A review of feature extraction methods on machine learning // *Journal Information System and Technology Management*. – 2021. – V. 6. – №. 22. – Pp. 51-59.

[9] Duo, A., Basagoiti, R., Arrazola, P. J., Aperribay, J., & Cuesta, M. The capacity of statistical features extracted from multiple signals to predict tool wear in the drilling process // *International Journal of Advanced Manufacturing Technology*. – 2019. – V.10 – Pp. 2133-2146.

[10] [https:// info. geology. gov. kz/ ru/ informatsiya/ spravochnik - mestorozhdenij - kazakhstan/](https://info.geology.gov.kz/ru/informatsiya/spravochnik-mestorozhdenij-kazakhstan/) (дата обращения 05 декабря 2023 года).

## REFERENCES\*

[1] Persaud, K.C.; Dodd, G. Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose // *Nature*. – 1982. – V. 299. – Pp. 352-355.



- [2] Karakaya D., Ulucan O., Turkan M. Electronic nose and its applications: A survey // International journal of Automation and Computing. – 2020. – V. 17. – №. 2. – Pp. 179-209.
- [3] Zaytsev V., Fedorov, F. S., Satybaldina D. et al. Rapid and accurate quality assessment method of recycled food plastics VOCs by electronic nose based on Al-doped zinc oxide // Journal of Cleaner Production. – 2023. – V. 418. – Paper N. 138042. <https://doi.org/10.1016/j.jclepro.2023.138042>
- [4] Satybaldina, D., Baydeldinov, M., Issainova, A., Alseitov O., Konyrkhanova A., Akhmetova, Z., Seilov, S. Artificial Olfactory System for Distinguishing Oil-Contaminated Soils // WSEAS Transactions on Environment and Development. – 2023. V. 19. – Pp. 951–960.
- [5] Satybaldina, D., Isajnova, A., Al'seitov, O., Bajdel'dinov, M., Seilov, Sh. Sozdanie gazoanaliticheskoy mul'tisensornoj sistemy tipa «jelektronnyj nos» i issledovanie ejo proizvoditel'nosti dlja zadach jekologicheskogo monitoringa // Vestnik KazATK. – 2023. – № 125(2). – S. 315–323.
- [6] Satybaldina D., Kalymova K., Sydykov D. Primenenie transfera obuchenija nejronnyh setej pri klassifikacii izobrazhenij // Vestnik KazNPU im. Abaja, serija «Fiziko-matematicheskie nauki». -2023. - №1 (81). - S.159-169.
- [7] TIBCO Data Science – Statistica. Available online: <https://www.tibco.com/products/data-science/> (data obrashhenija 05 dekabrja 2023 goda).
- [8] Suhaidi M., Kadir R. A., Tiun S. A review of feature extraction methods on machine learning // Journal Information System and Technology Management. – 2021. – V. 6. – №. 22. – Pp. 51-59.
- [9] Duo, A., Basagoiti, R., Arrazola, P. J., Aperribay, J., & Cuesta, M. The capacity of statistical features extracted from multiple signals to predict tool wear in the drilling process // International Journal of Advanced Manufacturing Technology. – 2019. – V.10 – Pp. 2133-2146.
- [10] [https:// info. geology. gov. kz/ ru/ informatsiya/ spravochnik - mestorozhdenij-kazakhstan/](https://info.geology.gov.kz/ru/informatsiya/spravochnik-mestorozhdenij-kazakhstan/) (data obrashhenija 05 dekabrja 2023 goda).

**Дина Сатыбалдина**, ф.м.ғ.к доцент, Л.Н.Гумилев атындағы Евразиялық ұлттық университеті, Астана, Қазақстан, [satybaldina\\_dzh@enu.kz](mailto:satybaldina_dzh@enu.kz)

**Алия Исаинова**, аға оқытушы, Л.Н.Гумилев атындағы Евразиялық ұлттық университеті, Астана, Қазақстан, [issainova\\_an@enu.kz](mailto:issainova_an@enu.kz)

**Наталья Глазырина**, PhD, доцент м.а., Л.Н.Гумилев атындағы Евразиялық ұлттық университеті, Астана, Қазақстан, [glazyrina\\_ns\\_1@enu.kz](mailto:glazyrina_ns_1@enu.kz)

**Алексей Шагабудинов**, студент, Л.Н.Гумилев атындағы Евразиялық ұлттық университеті, Астана, Қазақстан, [ashagabudinov@gmail.com](mailto:ashagabudinov@gmail.com)

## **КӨПСЕНСОРЛЫ ГАЗ-АНАЛИТИКАЛЫҚ ЖҮЙЕНІ ОҚЫТУ ҮШІН БЕЛГІЛЕРДІ АЛУДЫҢ СТАТИСТИКАЛЫҚ ӘДІСТЕРІ**

**Аңдатпа.** «Электронды мұрын» типті мультисенсорлы газ-аналитикалық жүйелер зиянды газдардың ағып кетуін анықтау, қауіпті заттарды анықтау, тамақ өнімдерінің сапасын бақылау және медициналық диагностикада қолданылады. Мұндай жүйелерде зерттелетін затқа сенсорлық жауаптардың жақындығына және оқыту кезеңінде пайдаланылатын ақпараттық белгілерге негізделген газды тануды жүзеге асыру үшін машиналық оқыту алгоритмдерін пайдаланады. Иіс классификациясының дәлдігін арттыру үшін ақпараттық белгілердің жиынтығын кеңейту үшін сенсорлық реакцияларды талдау әдістерін әзірлеу қажет. Жұмыста шикі мұнай үлгілерінен ауада ұшпа органикалық қосылыстардың болуына реакция ретінде сегіз «электрондық мұрын» датчиктерінің

электрлік сипаттамаларының өзгеруі туралы тәжірибелік деректерді статистикалық өңдеу нәтижелері берілген. Сипаттамалы статистикалық сипаттамаларды талдау, регрессия және кластерлік талдау әдістері корреляциялық факторларды таңдауға және олардың сенсорлық реакцияларға әсер ету дәрежесіне қарай дәрежелі факторларды таңдауға мүмкіндік беретіні көрсетілген.

**Түйінді сөздер.** Газ сенсоры, жасанды иіс сезу; иіс классификациясы, корреляциялық талдау, машиналық оқыту, сипаттамалық статистика, регрессиялық талдау.

**Dina Satybalдина**, candidate of physical and mathematical sciences, associate professor, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, satybalдина\_dzh@enu.kz

**Aliya Issainova**, senior lecturer, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, issainova\_an@enu.kz

**Natalya Glazyrina**, PhD, acting associate professor, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, glazyrina\_ns\_1@enu.kz

**Alexey Shagabudinov**, student, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, ashagabudinov@gmail.com

## STATISTICAL METHODS OF FEATURE EXTRACTION FOR TRAINING A MULTISENSOR GAS ANALYTICAL SYSTEM

**Annotation.** Multisensor gas analytical systems of the “electronic nose” type is used to detect leaks of harmful gases, detect hazardous substances, control the quality of food products, and in medical diagnostics. Such systems use machine learning algorithms to implement gas recognition based on the proximity of sensory responses to the analytes under study and information features used during the training phase. To increase the accuracy of odor classification, it is necessary to develop methods for analyzing sensory responses to expand the sets of information features. The results of statistical processing of experimental data on changes in the electrical characteristics of eight “electronic nose” sensors as reactions to the presence of volatile organic compounds in the air from crude oil samples are presented in the paper. It is shown that the analysis of descriptive statistical characteristics, regression and cluster analysis methods make it possible to select correlating factors and rank factors according to the degree of their influence on sensory responses.

**Keywords:** gas sensor, feature extraction, artificial olfaction, odor classification, correlation analysis, machine learning; descriptive statistics, regression analysis.

\*\*\*\*\*