Artificial Olfactory System for Distinguishing Oil-Contaminated Soils

DINA SATYBALDINA, MARAT BAYDELDINOV, ALIYA ISSAINOVA, OLZHAS ALSEITOV, ASSEM KONYRKHANOVA, ZHANAR AKHMETOVA, SHAKHMARAN SEILOV
Department of Information Security,
L.N. Gumilyov Eurasian National University,
2, Satpayev str., Astana, 010008,
REPUBLIC of KAZAKHSTAN

Abstract: - Oil-contaminated soils are a major environmental problem for Kazakhstan. Oil spills or leaks lead to profound changes in the physical and agrochemical properties of the soil and the accumulation of hazardous substances. Whilst there are many remote sensing techniques and complex laboratory methods for oil spill detection, developing simple, reliable, and inexpensive tools for detecting the presence of pollutants in the soil is a relevant research task. The study aims to research the possibilities of an electronic nose combining a chemical sensor array with pattern recognition techniques to distinguish volatile organic compounds from several types of hydrocarbon soil pollutants. An electronic nose system was assembled in our laboratory. It includes eight gas metal oxide sensors, a humidity and temperature sensor, an analog-digital processing unit, and a data communication unit. We measured changes in the electrical conductivity of sensors in the presence of volatile organic compounds released from oil and petroleum products and samples of contaminated and uncontaminated soils. The list of experimental samples includes six types of soils corresponding to different soil zones of Kazakhstan, crude oil from three oil fields in Kazakhstan, and five types of locally produced fuel oil (including gasoline, kerosene, diesel fuel, engine oil, and used engine oil). We used principal component analysis to statistically process multidimensional sensor data, feature extraction, and collect the volatile fingerprint dataset. Pattern recognition using machine learning algorithms made it possible to classify digital fingerprints of samples with an average accuracy of about 92%. The study results show that electronic nose sensors are sensitive to soil hydrocarbon content. The proposed approach based on machine olfaction is a fast, accurate, and inexpensive method for detecting oil spills and leaks, and it can complement remote sensing methods based on computer vision.

Key-Words: - artificial olfaction, crude oil, electronic nose, environment, machine learning, petroleum-derived products, pollution, sensor, soil, volatile organic compounds.

Received: April 23, 2023. Revised: July 11, 2023. Accepted: September 8, 2023. Published: September 27, 2023.

1 Introduction

Oil spills are a global problem; natural, intentional, or accidental oil spills can occur all around us. Significant oil pollution of the environment occurs in the territories adjacent to the sites of exploration, development, and operation of hydrocarbon deposits. Leaks from oil pipelines, spills when pumping oil to sea vessels, or accidents during oil transportation are also severe problems. One of the most recent significant cases is an environmental disaster on the Brazilian coast caused by the spill of about 2.5 million tons of Venezuelan oil from a ship that (intentionally or accidentally) dumped oil 700 km off the coast of Brazil, [1]. Finally, many standard fuels are refined petroleum products, and oil spills can occur during transportation, use, and disposal in many places, including residential areas.

Oil and petroleum products pollution is everywhere: in the soil layer, hydrosphere, and atmosphere due to the high level of volatile organic compounds emitted by spilled oil into the air, [2]. Oil introduces diverse chemical compounds into soil, water, and air, disrupting the established biogeochemical balance in ecosystems, [3].

Kazakhstan is the largest landlocked country in Asia, and the country ranks 9th in the world in terms of area. About 150 oil and 40 gas condensate fields are being developed in Kazakhstan, [4]. Kazakhstan has 30 billion barrels of proven oil reserves (12th rank in the world), and our country has increased oil production by 3.5 times over the past 30 years, [5].

At the same time, the development of the oil and gas industry leads to contamination of soil, water resources, and the atmosphere. Soil pollution in oil production areas is becoming increasingly significant, [6]. For example, 0.6 million hectares of soil were detected as contaminated with oil in western Kazakhstan, [7]. This region of the country has the highest density of oil fields. Crude oil spills disrupt soil structure and composition and reduce plant nutrient availability and uptake, [3]. The soil accumulates hydrocarbons and harmful microelements. which inhibit the growth of plants and microorganisms. The accumulation of aromatic hydrocarbons in plant cells increases human diseases, causing malignant tumors, [7].

Precise and rapid detection of crude oil and petroleum-derived products (PDPs) are beneficial to identifying the source and type of oil hydrocarbons, accurately estimating oil spread areas, evaluating the hazard level, and developing a response and recovery treatment to reduce environmental effects.

Aircraft satellites' Remote Sensing (RS) tools have been proven in oil spill detection and monitoring, [8]. Optical and microwave sensors, digital processing techniques, and pattern recognition were used for image classification and oil spill trajectory prediction, [9], [10]. The authors of the work, [10], showed that the admixture of nutrients impairs the accuracy of sensors when recognizing images of oil spills. New approaches for quickly and accurately classifying oil-contaminated soils can use alternative ones that complement remote sensing techniques based on computer vision.

Crude oil contains certain hydrocarbons (HCs) that have low boiling points and are classified as volatile organic compounds (VOCs) with known hazardous effects on human health and the air ecosystem, [2].

This context has prompted researchers to use artificial olfactory systems that mimic the mammalian olfactory system (an electronic nose, e-nose) and can discriminate odors by comparing their smells to previously studied patterns, [11]. The results of, [12], showed that the e-nose can differentiate soil contamination due to gasoline and diesel fuel leaks. The research in, [13], is another example of the use of an artificial olfactory system. The proposed approach made it possible to detect petroleum products adsorbed on different surfaces.

In this work, we investigated whether e-noses detect soil contamination and distinguish between some types of soils, oil, and PDP pollutants In this work, we investigated the performance of machine olfactory systems for classifying soil types and soil contaminants. A series of experiments were conducted to measure the response of electronic nose sensors to VOCs from all samples, including six uncontaminated soil samples, crude oil from three Kazakhstan oil fields, five PDPs from local producers (gasoline, kerosene, diesel fuel, engine oil and used engine oil), and soil samples with introduced petroleum pollutant. Then, the principal component analysis (PCA) was used for feature extraction and collection of the volatile fingerprint dataset. We evaluated the performance of decision trees and k-nearest neighbor classifiers using machine learning metrics. We demonstrated the high sensitivity of the sensors and the selective discrimination of VOC patterns of oil and PDP samples depending on the type, nature, and relative humidity of contaminated soils.

The rest of the paper consists of the following sections. In section 2 we described the materials and methods (materials, a description of the multisensory e-nose system, datasets, research tools, and research process). The experiment results on measuring sensory responses to the presence of petroleum and PDPs in soil samples, processing sensory data, features extracting, and performance evaluating by the machine learning algorithms are discussed in Section 3. The conclusion and future research are presented in Section 4.

2 Material and Methods

2.1 Samples

2.1.1 Soils

The relief forms of Kazakhstan are diverse (highlands, forest-steppe, steppe, desert-steppe cover, and desert), and it makes significant differences between the types of soils found in different geographical zones of the country. Chernozems, meadow-chernozem soils, saline soils, chestnut soils, and brown and grey-brown desert soils represent the diversity of soil types in Kazakhstan, [14].

Six soil types (chernozem, sand, birch grove soil, clay, slaked lime, and peat) were selected for laboratory experiments to cover a wide range of commonly encountered soils in Kazakhstan. The soil samples were collected locally from the surface layers (at a depth of 5–20 cm) away from roads and industries.

2.1.2 Contaminant

Crude oil and petroleum products spills release VOCs that readily evaporate into the air and source odors. The electronic nose can detect the effect of even slight changes in the relative amount of chemicals in the samples, [11]. Therefore, crude oil from several fields and five types of PDPs produced from different oil fractions with non-equal amounts of VOCs were chosen as samples (Table 1).

Table 1. List of Samples

Petroleum or PDP sample type	Oil field\ Producing company or brand of sample	Sample's physicochemical characteristics
Crude Oil	Alibekmola\ Kazakhoil Aktobe LLP	Light (0.835 g/cm^3), sour (about 2%) oil
Crude Oil	Alibek Yuzhny \ Joint Stock Company "Caspi Neft TME"	Light (0.842 g/cm^3), sour (up to 1.33%) oil
Crude Oil	Kardasyn North \ Sagiz Petroleum Company LLP	Heavy (0.945 g/cm^3), sweet (sour 0.2%) oil
Gasoline	Atyrau Oil Refinery LLP \ A-92-K5 with 92 octane number	Gasoline is produced from the light fraction of petroleum and contains hydrocarbons from four to twelve carbon atoms per molecule.
Diesel Fuel	Atyrau Refinery LLP \ DT-A-K5 Altay-45	Diesel Fuel is produced from the fraction of crude oil next in density to the gasoline fraction. Diesel fuel, like gasoline, contains a significant amount of volatile hydrocarbons with a strong odor. It contains hydrocarbons from nine to twelve five carbon atoms per molecule.
Kerosene	PetroKazakhstan Oil Products LLP \ RT	Kerosene is a mixture of liquid hydrocarbons containing eight to fifteen carbons in the hydrocarbon chain. Its density is higher than gasoline due to paraffin.
Engine Oil	High Industrial Lubricants & Liquids (HILL) Corporation \ «HILL Universal», SAE 5W-30	Engine Oil is produced from the dense hydrocarbon fraction of crude oil. It also contains some additives. Motor oils have a slight odor because they consist of non-light hydrocarbons with more than thirty carbon atoms per molecule.
Used Engine Oil	High Industrial Lubricants & Liquids (HILL) Corporation \ Used «HILL Universal», SAE 5W-30	Used Engine Oil may change their composition during use and it may potentially contain more harmful contaminants than unused engine oil.

Oil products used for mechanized land cultivation are also a source of soil pollution, [15], [16]. The study used two types of engine oil: a new synthetic engine oil manufactured by High Industrial Lubricants & Liquids (HILL) Corporation and used engine oil (UMO) of the same brand collected from a local garage.

2.2 E-nose

Our artificial olfactory system consists of a multisensory array, an analog-to-digital conversion unit, and a digital data transmission unit. MQ series sensors from Hanwei Electronics were used as gas sensors. Sensor color codes and target gases are presented in the legends of Figure 1 (Appendix), Figure 2 (Appendix) and Figure 3 (Appendix). These sensors belong to the class of metal oxide sensors.

The choice of this type of gas sensor is associated with its low cost, small size, and relatively low power consumption. The MOS sensors use a two-electrode system in which the sensitive layer of tin dioxide (SnO_2) has a variable resistance depending on the concentration and type of gases or gas mixtures being studied. The interaction of the target gas with the sensitive layer occurs through a reversible redox reaction, as a result of which the electrical properties of the SnO_2 layer change. These electrical properties are translated into measurable parameters such as sensor resistances or voltage across a load resistor in a sensor circuit.

MOS sensors, SHT75 temperature, and humidity sensors were grouped on a separate board and connected via an analog switch to an analog-to-digital converter (ADC) on the Arduino Nano module. The ADC has a 16-bit resolution and converts input voltages from 0 to 5 volts into integers from 0 to 65535 (2^{16}-1). The digitized data is stored on a Raspberry Pi 4V microcomputer, which has built-in support for 5G Wi-Fi and Bluetooth 5. In the future, it is planned to implement an intelligent sensor data processing unit on this budget single-board computer.

MOS sensors need to be heated. Arduino cannot provide enough power. Therefore, the system uses an external power source connected via the USB port. Sensors, a microcontroller, a microcomputer, a power supply circuit, and a forced air supply unit are placed inside a metal box.

2.3 Data Collection

Datasets were collected in a sequential series of experiments measuring sensory responses from:

– uncontaminated soil samples,

– samples of Contaminant (crude oil and PDPs),
– soil samples with introduced pollutants.

Soil samples of the same mass (70 grams), liquid samples of oil and PDPs (volume 100 μl), and mixtures of soils and pollutants were placed in a glass vial connected to the nasal measuring chamber.

Each experiment was repeated three times. The output signal from the array of sensors, U_{ADC}, was measured at a sampling frequency of 1 Hz during the total time of the experiment for each sample (the air was measured in the first part of the experiment, then the sample was introduced, and the flow was switched to air for 30 min in the final part). Experimental data were recorded in separate files (txt format). Each file consists of more than 7000 lines with ten attributes, namely data year-month-day and time (hours-minutes-seconds), responses from 8 sensors, temperature (degrees Celsius), and humidity (%). The dataset contains all the data and significant outliers.

The total number of records exceeded 170,000. The data set was divided into two parts - 80% was used as a training set and 20% of the data was used at the classifier testing stage. In turn, part of the training sample (30%) was used for validation to control the training of machine learning models.

2.4 Data Analysis

Pattern recognition methods distinguish between uncontaminated and contaminated soil samples, soil samples, and petroleum products. Preliminary feature extraction procedures are required to prepare datasets for input to recognition systems.

We used principal component analysis to statistically process measurements of sensory responses to experimental samples and extract the most significant features from multidimensional sensory data.

Vector representations of the selected features as volatile odor fingerprints from samples constituted databases for training and testing classifiers based on two machine learning algorithms (decision tree and k-nearest neighbor method), [17].

We use code written in Python 3.9.15 and Scikit-learn open source machine-learning package, [18]. The performance metrics (the confusion matrix, accuracy, precision, and f-score) are used, [10].

For interactive calculations and experimental data processing, a Python application has been developed (in ipynb format) containing source codes, input data, and calculation results in numerical and graphical representation.

3 Results and Discussion

3.1 E-nose Results

of Figure 1 (Appendix), Figure 2 (Appendix) and Figure 3 (Appendix) show examples of visualization of experimental results in the form of time-dependent electrical characteristics of sensors as sensor responses to VOCs from the samples.

The figures show that the sensors of the gas analysis system react differently to experimental samples of uncontaminated and contaminated soils, crude oil, and PDPs, which allows for obtaining volatile fingerprints.

An analysis of the experimental data indicates the presence of a correlation between the physicochemical properties of crude oils from three fields (Table 1) and the sensory responses to the petroleum samples (Figure 1, Appendix). Light oils contain more volatile hydrocarbons than denser oils. Accordingly, the level of sensory responses to the smell from VOCs is higher. The sulfur content in oil can also affect the shape of sensory responses.

Gasoline corresponds to the lightest oil fraction. This PDP has the wealthiest odor and fastest evaporating. Therefore, we observe high-amplitude sensory responses of the e-nose to the presence of VOCs in motor gasoline and a rapid drop in sensory response due to the high evaporation rate (Figure 2a, Appendix). Diesel fuel corresponds to the average fraction of oil, has fewer VOCs compared to gasoline, and the sensory responses of the electronic nose in the case of diesel fuel are less intense (Figure 2b, Appendix), its shapes indicate the average rate of evaporation and weathering.

Gasoline and kerosene have different chemical chain lengths, but both have strong odors. Kerosene contains paraffin and has a oilier structure. Therefore, the experimental electrical characteristics of MOS sensors in the presence of kerosene have a high intensity, as in the case of gasoline, and decrease more slowly (Figure 2c, Appendix).

Engine oils are thick mixtures of high molecular weight hydrocarbons obtained by distillation of heavy oil fractions (fuel oil and tar). The low amplitudes of sensory responses to motor oil (Figure 2d, Appendix) indicate lower concentrations of volatile organic compounds than lighter petroleum products (gasoline, diesel fuel, or kerosene). Remarkable results have been obtained with used engine oil. Sensor responses indicate the presence of a significant amount of VOCs in the used oil (Figure 2e, Appendix). The closeness of the used oil sensor data to the gasoline and diesel fuel VOC fingerprints indicates the possibility of:

- malfunctions in the engine that lead to the ingress of gasoline or diesel fuel into the motor oil;
- or an increase in toxic components in oil during use.

Figure 3 (Appendix) shows that the gas analysis system reacts differently to experimental samples of uncontaminated and contaminated soils, which allows them to differentiate.

3.2 Data Analysis Results

The work uses eight gas sensors and a temperature and humidity sensor. Sensory responses to the samples with different quantities of the studied substances and various temperature and air humidity values represent multidimensional spaces of dependent and independent features. Feature extraction procedures are necessary for fast and accurate classification of recognized objects, [17], [19], [20].

We use the statistical method of principal component analysis (PCA) to extract significant features from sensory data and form a digital fingerprint of each class of samples under study.

The first step of the method is data standardization, [21]. The normalized value is the ratio of the difference between the sensory response value and the mean value to the standard deviation value. Then we calculate the covariance matrix for the normalized data, [22]. Calculating the eigenvectors and eigenvalues allows us to select the principal components. Finally, we can express the standardized variables in terms of principal component scores, [23].

An example of visualization of the results of PCA analysis of sensory data is presented in Figure 4.

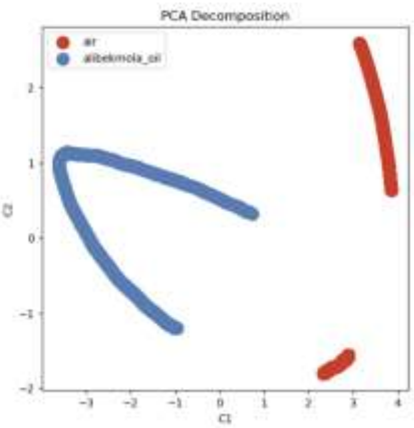

Fig. 4: Visualizing PCA results using Biplot for sensor responses to the air and the crude oil from the Alibekmola oil fields.

In Figure 4, each point represents sensor data from two classes (air and oil). Let us recall that the measurement of sensory responses for each sample was carried out in stages: in the first part of the experiment, the air was measured, then the sample was introduced, and in the final part, the flow was switched to stand for 30 min). It can be seen that points belonging to the same class (oil or air) are closer to each other, indicating that the features were correctly extracted for constructing digital volatility fingerprints.

Table 2 presents the performance estimates of the volatile fingerprint classifiers of the studied samples using two machine-learning algorithms. Pattern recognition using machine learning algorithms made it possible to classify digital fingerprints of samples with an average accuracy of about 92%.

Table 2 shows that the accuracy of digital fingerprint recognition is above 90% in most cases. The average classification accuracy of crude oil brands and types of petroleum products is 92.15% and 91.33% when using the decision tree and the KNN method, respectively.

Table 2. Performance Results of Pattern Recognition models to classify crude oils and PDPs samples on test stage

Petroleum or PDP sample type	**Performance metrics**	
	KNN	***Decision Trees***
Crude Oil Alibekmola	Accuracy: 97.76 % F1 SCORE: 97.72 % PRECISION: 98.06 % RECALL: 97.49 %	Accuracy: 98.32 % F1 SCORE: 98.29 % PRECISION: 98.53 % RECALL: 98.11 %
Crude Oil Alibek Yuzhny	Accuracy: 95.59 % F1 SCORE: 95.53% PRECISION: 96.2 % RECALL: 95.25 %	Accuracy: 95.97 % F1 SCORE: 95.92% PRECISION: 96.51 % RECALL: 95.65 %
Crude Oil Kardasyn North	Accuracy: 95.97 % F1 SCORE: 95.92% PRECISION: 96.51 % RECALL: 95.65 %	Accuracy: 96.43 % F1 SCORE: 96.23 % PRECISION: 97.19% RECALL: 95.55%
Motor Gasoline	Accuracy: 96.23 % F1 SCORE: 96.19% PRECISION: 96.71% RECALL: 95.93 %	Accuracy: 97.32 % F1 SCORE: 97.29 % PRECISION: 97.53 % RECALL: 97.11 %
Diesel Fuel	Accuracy: 85.69 % F1 SCORE: 85.47 % PRECISION: 85.64 % RECALL: 85.35 %	Accuracy: 87.8 % F1 SCORE: 87.57 % PRECISION: 87.9 % RECALL: 87.37 %
Engine Oil	Accuracy: 95.66 % F1 SCORE: 95.6 % PRECISION: 96.25 % RECALL: 95.31 %	Accuracy: 96.22 % F1 SCORE: 96.18 % PRECISION: 96.71 % RECALL: 95.93 %
Used Motor Oil	Accuracy: 72.45 % F1 SCORE: 72.1 % PRECISION: 76.31 % RECALL: 80.15 %	Accuracy: 72.96 % F1 SCORE: 72.58 % PRECISION: 76.54 % RECALL: 80.52 %

The decrease in classification accuracy for diesel fuel and used engineering oil is due to the proximity of sensory responses for these samples. Therefore, part of the data for diesel fuel supplied to the input of the classifiers is incorrectly recognized as volatility fingerprints of used motor oil, and vice versa.

4 Conclusion

In the present paper, we investigated artificially polluted soils to detect soil contamination and distinguish between some types of soil, oil, and PDP pollutants using alternative sensor technologies that can replace or supplement computer vision and RS techniques.

We examined six soil types found in different geographic areas of Kazakhstan, eight different pollutants, including crude oil from three Kazakhstan fields, and commercial gasoline, kerosene, diesel fuel, motor oil, and used motor oil. The sensor responses to volatile organic compounds of soil, crude oil, and petroleum product samples recorded by the electronic nose were processed using a statistical analysis method. After the feature extraction stage, feature vectors were used to train and test classifiers based on a machine learning algorithm. The experimental results showed that the artificial olfactory system is sensitive to different types of soil and the composition of petroleum products. A machine learning model implemented in Python recognizes contaminated and uncontaminated soils with high accuracy and the kind of oil and petroleum products. The proposed approach to detecting oil-contaminated soils based on inexpensive and compact devices such as an electronic nose is a good alternative to oil spills' current remote sensing methods.

The proposed approach will be used in future studies to determine the source and type of oil pollution on soil samples from oil production fields and other contaminated areas.

References:

[1] Marcelo de Oliveira Soares et al., Oil spill in South Atlantic (Brazil): Environmental and governmental disaster, *Marine Policy*, Vol. 115, 2020, 103879.

[2] Ana Cláudia Souza et al., Oil spills characteristics, detection, and recovery methods: A systematic risk-based view, *Journal of Loss Prevention in the Process Industries*, Vol. 80, 2022, 104912.

[3] Hussein, Z. S., Hamido, N., Hegazy, A. K., El-Dessouky, M. A., Mohamed, N. H., & Safwat, G. (2022). Phytoremediation of crude petroleum oil pollution: a review, *Egypt. J. Bot.*, Vol. 62, No.3, 2022, pp.611-640.

[4] Bealessio B., Alonso N., Mendes N., Sande A., & Hascakir B., A review of enhanced oil recovery (EOR) methods applied in Kazakhstan, *Petroleum*, Vol. 7 (1), 2021, pp.1–9.

[5] BP Statistical Review of World Energy, accessed 8 July 2021, 70th Edition https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2021-full-report.pdf.

[6] Alimbaev, T., Yermagambetova, K., Kabyltayeva, S., Issayev, A., Kairat, Z., & Mazhitova, Z., Environmental problems of the oil and gas industry in Kazakhstan, *E3S Web of Conferences*, Vol. 215, 2020, p.03008.

[7] Akhmetov L.I. et al., Recent Advances in creating biopreparations to fight oil spills in soil ecosystems in the sharply continental climate of the Republic of Kazakhstan, *Processes*, Vol.10, No.3, 2022, 549.

[8] Fingas M., Brown C., & Daniels J., An update on oil spill remote sensing, *Advances in Environmental Research*, Vol.80, 2021, pp.95-137.

[9] Shamsudeen T., Advances in remote sensing technology, machine learning and deep learning for marine oil spill detection, prediction and vulnerability assessment, *Remote Sensing*, Vol. 12, No. 20, 2020, p.3416.

[10] Al-Ruzouq R. et al., Sensors, features, and machine learning for oil spill detection and monitoring: A review, *Remote Sensing*, Vol.12, No.20, 2020, p.3338.

[11] Zaytsev V. 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*, Vol. 418, 2023, 138042.

[12] Bieganowski A. et al., Evaluation of hydrocarbon soil pollution using e-nose, *Sensors*, Vol.18, No.8, 2018, 2463.

[13] Ferreiro-González, M., Barbero, G. F., Palma, M., Ayuso, J., Álvarez, J. A., & Barroso, C. G., Characterization and differentiation of petroleum-derived products by E-nose fingerprints, *Sensors*, Vol.17, No.11, 2017, 2544.

[14] Pachikin, K., Erokhina, O., & Funakawa, S., Soils of Kazakhstan, their distribution and mapping, in *Novel measurement and assessment tools for monitoring and management of land and water resources in*

agricultural landscapes of Central Asia, Springer International Publishing, Switzerland, 2014, pp.519-533.

[15] Vazquez-Duhalt R., Environmental impact of used motor oil, *Science of the total environment*, Vol.79.1,1989, pp.1–23.

[16] Mishra A., Siddiqi H., Kumari U., Behera I., Mukherjee S. & Meikap B., Pyrolysis of waste lubricating oil/waste motor oil to generate high-grade fuel oil: A comprehensive review, *Renewable and Sustainable Energy Reviews*, Vol. 150, 2021, 111446.

[17] Isah Salim Ahmad, Shuai Zhang, Sani Saminu, Lingyue Wang, Abd El Kader Isselmou, Ziliang Cai, Imran Javaid, Souha Kamhi, Ummay Kulsum, Deep Learning Based on CNN for Emotion Recognition Using EEG Signal, *WSEAS Transactions on Signal Processing*, vol. 17, 2021, pp.28-40.

[18] Hao J., & Ho T., Machine learning made easy: a review of a scikit-learn package in a python programming language, *Journal of Educational and Behavioral Statistics*, vol.44(3), pp.348-361, March 2019.

[19] Md. Salah Uddin Yusuf, Azmol Ahmed Fuad, Real Time Implementation of Face Recognition based Smart Attendance System, *WSEAS Transactions on Signal Processing*, vol. 17, 2021, pp.46-56.

[20] Meili Dai, Intelligent Correction System of Students’ English Pronunciation Errors Based on Speech Recognition Technology, *WSEAS Transactions on Advances in Engineering Education*, vol. 18, 2021, pp.192-198.

[21] Jackson J. E., *A user's guide to principal components*, Toronto, Canada: John Wiley & Sons Inc., 2005.

[22] Fadda ML, Moussaoui A. Faults Detection and Diagnosis Approach Using PCA and SOM Algorithm in PMSG-WT System, *WSEAS Transactions on Systems*, Vol. 16, 2017, pp.159-67.

[23] Radhia Fazai, Okba Taouali, Nasreddine Bouguila, Adaptive Fault Detection Methods Based on PCA Technique, *WSEAS Transactions on Circuits and Systems*, vol. 14, 2015, pp.337-346.

APPENDIX

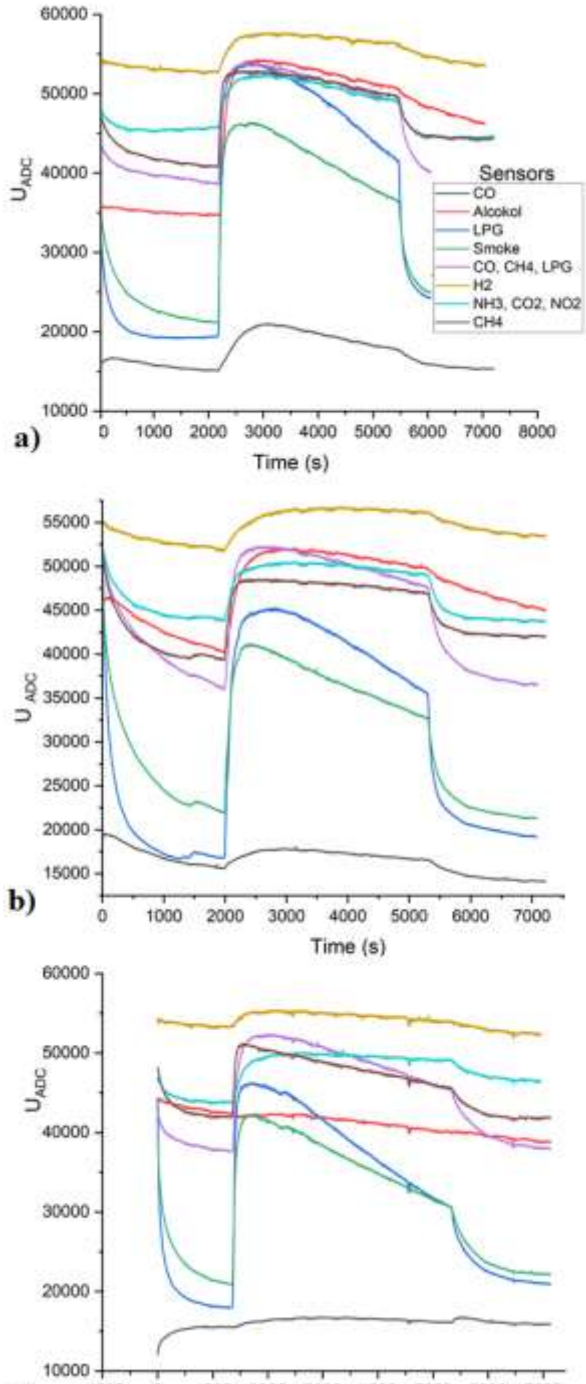

Fig. 1: The time-dependent electrical characteristics of e-nose sensors for crude oil from three Kazakhstan oil fields: (a) Alibekmola, (b) Alibek Yuzhny, (c) Kardasyn North.

Dina Satybaldina, Marat Baydeldinov,
Aliya Issainova, Olzhas Alseitov,
Assem Konyrkhanova, Zhanar Akhmetova,
Shakhmaran Seilov

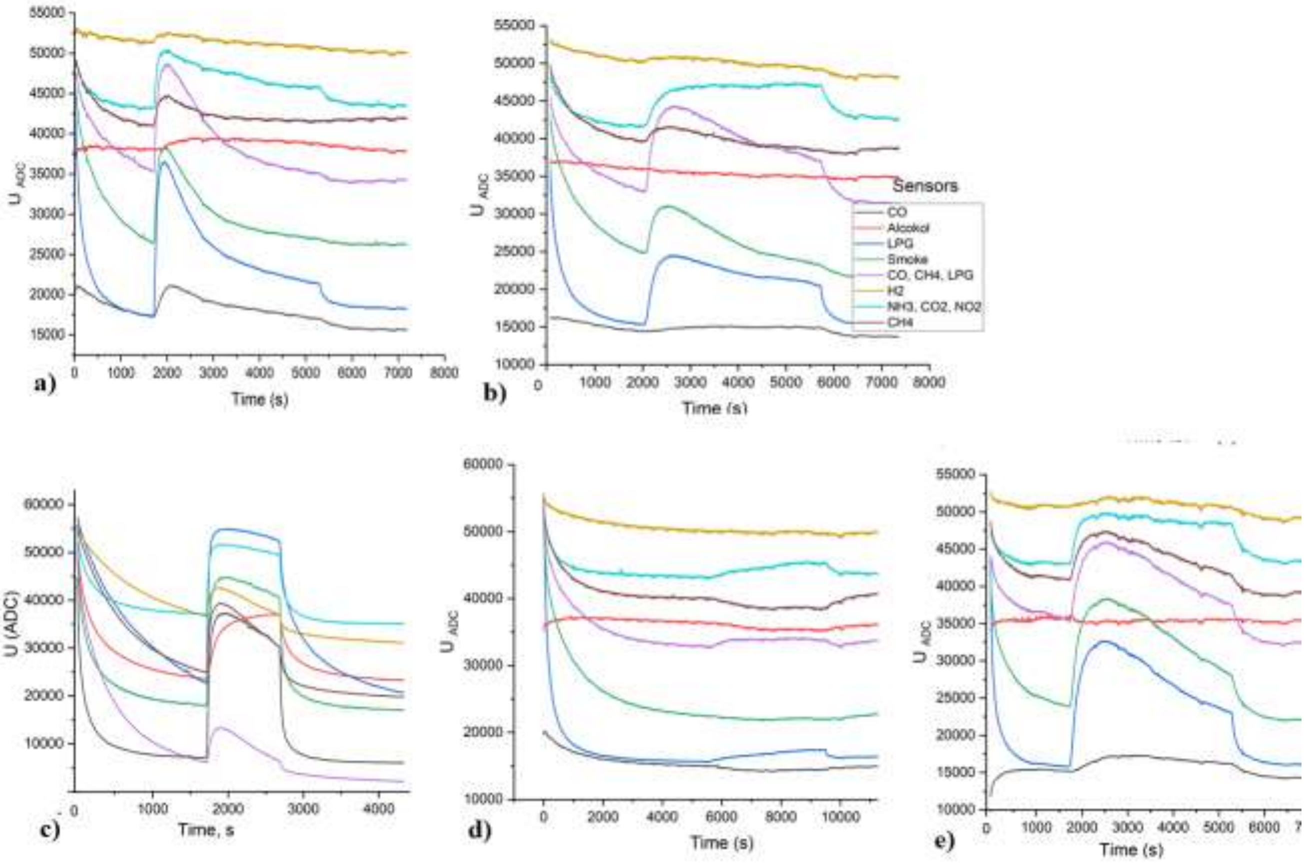

Fig. 2: The time-dependent electrical characteristics of e-nose sensors for PDPs: (a) Gasoline; (b) Diesel fuel; (c) Kerosene; d) Engine oil; (e) Used engine oil.

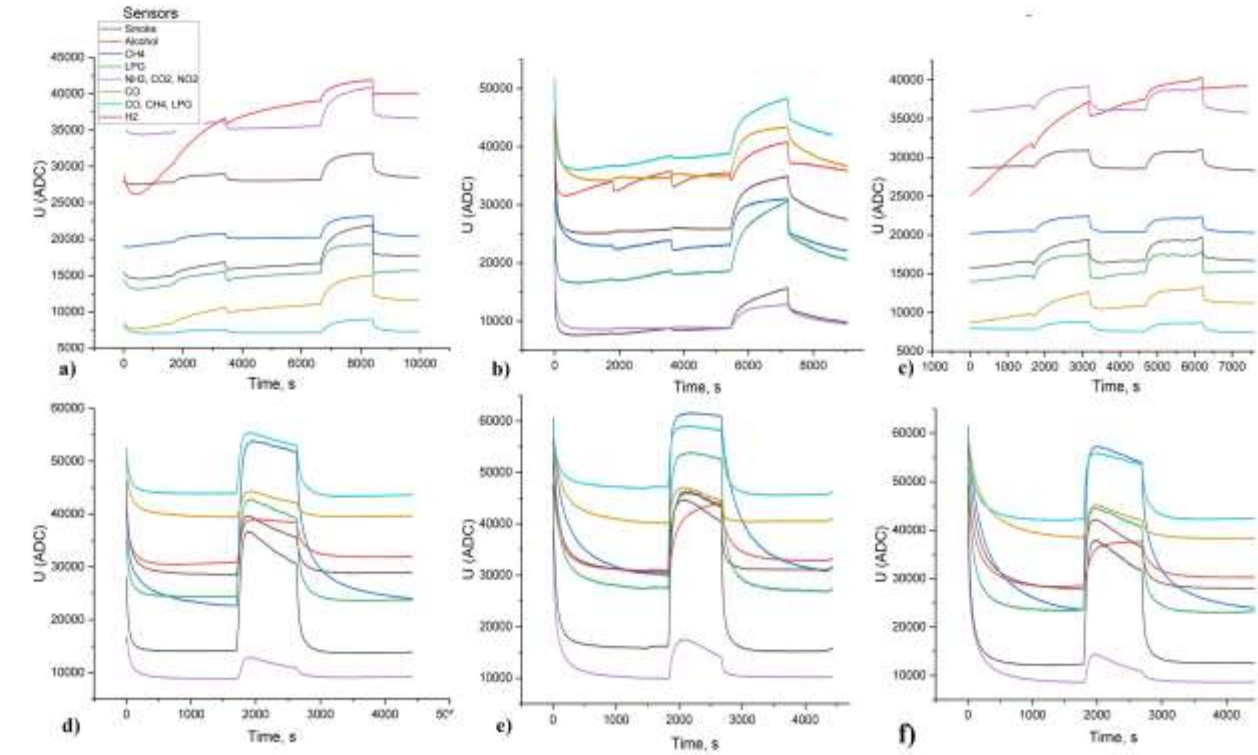

Fig. 3: The time-dependent electrical characteristics of e-nose sensors for the uncontaminated and contaminated soil samples: (a) chernozem (dry and wet), (b) sand (dry and wet), (c) birch grove soil (dry and wet), d) chernozem sand + kerosene; (e) sand + Kerosene, (f) birch grove soil + kerosene.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Dina Satybaldina: Supervision, Conceptualization, Methodology, Validation, Investigation, Writing – original draft.
- Marat Baydeldinov: Experimental Methodology, Investigation, Writing – original draft.
- Aliya Issainova, PhD student: Experiments execution, Data curation, ML models implementation in Python, Data analysis.
- Olzhas Alseitov, Master student: Experiments execution, Data curation.
- Assem Konyrkhanova: Methodology, Data Analysis, Writing – original draft.
- Zhanar Akhmetova: Methodology, Investigation, Validation.
- Shakhmaran Seilov: Project supervising, Resources.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP14872171).

Conflict of Interest

The authors have no conflict of interest to declare.