UDC 528.8:528.7:528.4:528.06(045)
DOI:10.18372/1990-5548.84.20199

[1]I. G. Prokopenko,
[2]S. I. Alpert,
[3]M. I. Alpert,
[4]A. Y. Dmytruk

CLASSIFICATION OF SENTINEL-2 IMAGERY USING RAYLEIGH DISTRIBUTION MODELING

[1,4]Faculty of Aeronautics, Electronics and Telecommunications, State University "Kyiv Aviation Institute", Kyiv, Ukraine
[2]Architecture, Construction and Design Faculty, State University "Kyiv Aviation Institute", Kyiv, Ukraine
[3]The Department of Information Systems and Technologies, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
E-mails: [1]igorprok48@gmail.com ORCID 0000-0003-4169-3774,
[2]sonyasonet87@gmail.com ORCID 0000-0002-7284-6502,
[3]max292009@gmail.com ORCID 0000-0002-8938-1473,
[4]gyhyre@gmail.com ORCID 0000-0002-6410-4680

Abstract**—***Nowadays land cover classification from satellite imagery is one of most actual and important problems in remote sensing. Multispectral satellite images such as Sentinel-2 images provide high-resolution imagery in different spectral bands, enabling detailed distinguishing of surface objects. This study presents a method of multispectral satellite image classification based on Rayleigh distribution, maximum likelihood method and likelihood functions. It was considered three land cover classes, such as "Water", "Vegetation", and "Buildings", applying three spectral bands (Red spectral band, Green spectral band and Blue spectral band). Proposed classification procedure includes modeling spectral distributions with the Rayleigh probability distribution. The Rayleigh distribution parameters for each class and each spectral band are estimated from training data via the proposed formula. The ESA SNAP software is applied for image processing. Maximum likelihood method is applied for classification procedure. In remote sensing this method is used to classify pixels in satellite imagery into different classes. This method is based on assigning each pixel to the class, for which has the highest probability of belonging. It was described the methodology, including data preparation using the ESA SNAP software and data analysis in Microsoft Excel. The mathematical formulation of the Rayleigh distribution and the mathematical algorithm of calculation of likelihood functions for each class and for each spectral band have been considered. Results include fitted Rayleigh distribution parameters for each class and for each spectral band, classification maps, calculation of likelihood functions and classification result. The classification result depends on which class the maximum likelihood function corresponds to. An example has been considered where the class "Vegetation" is determined using the maximum likelihood method and Rayleigh distribution. The proposed approach can be applied for land-cover classification, ecological monitoring, agriculture and geological tasks.*

Keywords—Likelihood functions; maximum likelihood method; Rayleigh distribution; remote sensing; satellite image classification; signal processing of spectral bands of satellite images**.**

I. INTRODUCTION

Accurate land cover classification from satellite imagery is a fundamental task in remote sensing, supporting environmental monitoring, urban planning, and resource management. Multispectral optical satellites such as the European Space Agency's Sentinel-2 provide high-resolution imagery in several spectral bands, enabling detailed distinguishing of surface materials. The Sentinel-2 MultiSpectral Instrument (MSI) captures imagery in different spectral bands, including visible Red, Green and Blue bands at 10-meter spatial resolution. These visible bands are particularly useful for distinguishing surface cover types like water bodies, vegetation, and urban or buildings due to their different reflectance characteristics. Nowadays a lot of various image classification methods are known [7], [8], [9]. Supervised classification approaches rely on statistical models of the spectral signatures of known classes. One of the most popular methods is the Maximum Likelihood Classification (MLC). This classifier is applied to classify pixels in satellite imagery into different classes. The Maximum Likelihood Classification is based on assigning each pixel to the class that has the highest probability of producing the observed pixel values, based on the statistical characteristics of the classes. Each pixel is assigned to the class for which has the highest probability of belonging. In a classical MLC, it is assumed that the distribution of reflectance values for each class in each spectral band is approximately normal (Gaussian). However, it has been noted some disadvantages of the normal distribution by Mauricio Acuna, Geonhwi Jung, Joowon Park, Bruna G. Palm, Balakrishnan N. and others [1], [3]. The normality assumption may not always hold, especially for reflectance data that are non-negative and often imprecise. That's why in this paper, we explore an alternative probability distribution – the Rayleigh distribution – to model the pixel intensity histograms for each class and each spectral band. The Rayleigh distribution is a continuous probability distribution defined only for nonnegative values. It is applied to describe the distribution of pixel amplitudes in radar images. The Rayleigh distribution is used for classification of homogeneous regions [2], [3]. In the context of optical imagery, class-specific reflectance histograms may be described by Rayleigh distribution. The **goal** of this study is to classify each pixel of a Sentinel-2 image into one of three categories, such as: "Water", "Vegetation", or "Buildings", applying Rayleigh distribution and likelihood functions. We detail the methodology, including how training data are obtained and normalized, how the Rayleigh distribution parameters are estimated, and how the likelihood functions are applied. We also describe the software tools used: the Sentinel Application Platform (SNAP) for image processing and sample extraction, and Microsoft Excel for histogram analysis and calculations of likelihood functions. SNAP is an open-source ESA software platform that provides a common architecture for processing Sentinel satellite data, making it ideal for tasks such as band extraction and selection of training samples (ROI selection). Excel has been chosen to perform statistical computations (computing histograms, fitting distribution parameters, and applying formulas).

II. Methodology

We used Sentinel-2 imagery covering an area of a part of Kyiv region (containing a mix of water, vegetated land, and buildings) for 25.08.2025 (Fig. 1). Specifically, the study area includes open water bodies (such as river), dense green vegetation (forest or cropland), and built-up urban features (buildings, roads). The Sentinel-2 data provides 10 m resolution in the visible bands, which allows identification of these classes at a fine spatial scale. The image was loaded and examined in SNAP software. In SNAP, we performed basic preprocessing steps: selecting the Red, Green, and Blue spectral bands. The digital pixel values in this product are physically scaled reflectances (surface reflectance) in each spectral band, typically given as dimensionless reflectance factor (with values between 0 and 1).

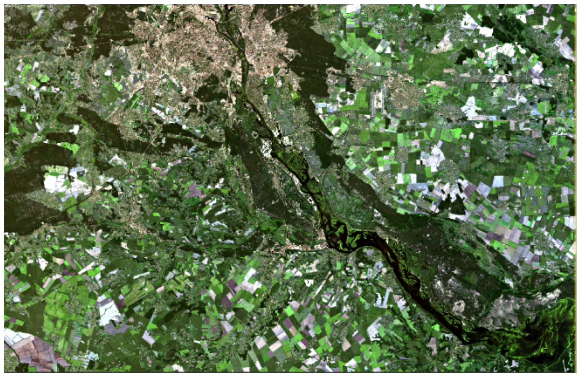

Fig. 1. Sentinel-2 satellite image of the part of Kyiv region (25.08.2021)

Then we define a Region of Interest and Training Samples (Fig. 2). In order to model the class distributions, we needed representative sample pixels for each of the three classes [4] – [6], [10].

Using SNAP's visualization and ROI tools, we identified three regions:

1) Water: an area entirely covering a section of a water body (for example, part of a river) was selected to gather water pixels.

2) Vegetation: an area of dense vegetation (forest or field) was selected for vegetation pixels.

Buildings: an urban area (with building rooftops, etc.) was selected for built-up class pixels.

Then we conduct a classification procedure and get a classification map (Fig. 3). Maximum Likelihood classifier is applied. The Maximum Likelihood Classification procedure is based on assigning each pixel to the class that has the highest value of the probability of producing the observed pixel values, based on the statistical characteristics of the classes. Each pixel is assigned to the class for which it has the highest probability of belonging [1].

Fig. 2. Training samples

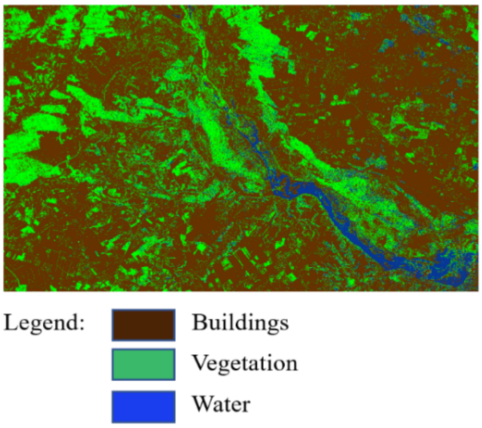

Fig. 3. Classification map

From each ROI, we extracted the pixel intensity values in the Red, Green and Blue bands. These intensities (reflectance values) served as the training data for estimating the statistical distribution of each class in each spectral band. We exported these values and imported them into Microsoft Excel for analysis. We also get histograms of the pixel intensities for each class and for each spectral band. For each class-band combination, we constructed a histogram of the pixel intensities. *X*-axis displays the numerical values of pixel measurements (signal intensity), and *Y*-axis displays the number of pixels that have the corresponding value on the X-axis. Let's note, that each histogram can be described by an empirical probability density function. For example, we constructed a histogram of the pixel intensities for class "Vegetation" for the Blue Band (Fig. 4).

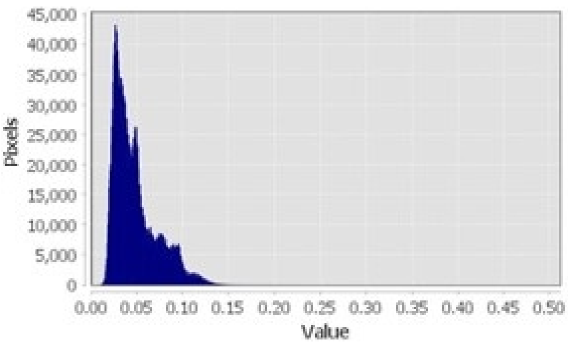

Fig. 4. Distribution of pixel intensities and their corresponding number for the class "Vegetation" and for the Blue Band

III. RAYLEIGH DISTRIBUTION MODELING

We hypothesize that the distribution of pixel intensities for each class in each spectral band can be modeled by a Rayleigh distribution. The Rayleigh distribution is defined by a single parameter, commonly denoted $\delta > 0$, which determines the spread of the distribution [3]. Let's note, that the Rayleigh distribution is a continuous probability distribution for nonnegative-valued random variables. The probability density function of the Rayleigh distribution is defined by the formula:

$$f(x;\delta) = \begin{cases} \dfrac{x}{\delta^2}\exp\left(-\dfrac{x^2}{2\delta^2}\right), & x \ge 0, \\ 0, & x < 0, \end{cases} \qquad (1)$$

where δ is the scale parameter of the distribution.

V. PARAMETER ESTIMATION

For each class and each spectral band, we estimate the Rayleigh parameter δ from the training pixel values via the formula:

$$\delta_k^{*2} = \frac{1}{2n}\sum_{i=1}^{n} x_i^2, \quad k = 1, \dots, K, \qquad (2)$$

where x_i are values of pixel intensities of training sample; K is the total number of classes; n is the number of values of pixel intensities of training sample.

Formula (2) substitute into the formula (1) and get formula (3), that describes the Rayleigh distribution with parameter δ_{1B}^{*}, that describes the graph of the distribution of pixel intensities for class "Vegetation" in Blue spectral band:

$$f_{1B}(x) = \begin{cases} \dfrac{x}{\delta_{1B}^{*2}}\exp\left(-\dfrac{x^2}{2\delta_{1B}^{*2}}\right), & x \ge 0, \\ 0, & x < 0, \end{cases} \qquad (3)$$

Similarly, we get the Rayleigh distributions with another 8 parameters, that describe the graphs of the distribution of pixel intensities for another 8 class-band combinations [3]. We applied the values of pixel intensities of the training sample x_i for the construction estimates of the scale parameter δ for 3 classes and for 3 spectral bands. Next, we need to create a new test sample that needs to be classified. In this same satellite image, we select another test sample for one specific class (for example, we select a test sample for a class "Vegetation") and conduct the classification procedure again using the maximum likelihood method. As a result, we get new pixel intensities for 3 classes and for 3 spectral bands. We also get 9 new histograms (graphs) that display the distribution of pixel intensities for all class-band combinations. We substitute new values of pixel intensities y_i, $i = 1, \dots, m'$ and parameter estimates δ_k^{*2} into the formulas for the densities of the Rayleigh distribution and calculate new probability density functions for the new test sample. Then we calculate the log-likelihood function for the new test sample for class "Vegetation" for 3 spectral bands (Blue Band, Red Band, Green Band). The logarithm of the likelihood function for the class "Vegetation" is equal to the sum of the logarithms of the partial likelihood functions for 3 spectral bands. It is defined by the next formulas (4) – (7):

$$\ln L_1(\delta_1^*) = \ln L_B(\delta_1^*) + \ln L_R(\delta_1^*) + \ln L_G(\delta_1^*), \quad (4)$$

where

$$L_B(\delta_1^*) = \prod_{i=1}^{m} f\left(y_{iB}, \delta_{1B}^*\right) = \prod_{i=1}^{m}\left(\frac{y_{iB}}{\delta_{1B}^{*2}}\exp\left(-\frac{y_{iB}^2}{\delta_{1B}^{*2}}\right)\right), \quad (5)$$

or

$$L_B(\delta_1^*) = \left(\frac{1}{\delta_{1B}^{*2}}\right)^m \left(\prod_{i=1}^{m} y_{iB}\right)\exp\left(-\frac{1}{2\delta_{1B}^{*2}}\sum_{i=1}^{m} y_{iB}^2\right). \quad (6)$$

Applying the properties of natural logarithms, we get formula:

$$\ln L_B(\delta_1^*) = -2m \cdot \ln \delta_{1B}^* + \sum_{i=1}^{m} y_{iB} - \frac{1}{2\delta_{1B}^{*2}}\sum_{i=1}^{m} y_{iB}^2. \quad (7)$$

So, we found formulas (8), (9) for the partial log-likelihood function of the Rayleigh distribution for the first class "Vegetation" and for the Blue Band. Similarly, we get partial log-likelihood functions of the Rayleigh distribution for the first class "Vegetation" and for the Red Band and Green Band accordingly:

$$\ln L_R(\delta_1^*) = -2m' \cdot \ln \delta_{1R}^* + \sum_{i=1}^{m'} y_{iR} - \frac{1}{2\delta_{1R}^{*2}}\sum_{i=1}^{m'} y_{iR}^2, \quad (8)$$

$$\ln L_G(\delta_1^*) = -2m'' \cdot \ln \delta_{1G}^* + \sum_{i=1}^{m''} y_{iG} - \frac{1}{2\delta_{1G}^{*2}}\sum_{i=1}^{m''} y_{iG}^2. \quad (9)$$

Similarly, we get a log-likelihood function for the second class "Water":

$$\ln L_2(\delta_2^*) = \ln L_B(\delta_2^*) + \ln L_R(\delta_2^*) + \ln L_G(\delta_2^*). \quad (10)$$

The log-likelihood function for the third class "Buildings" is defined by the formula:

$$\ln L_3(\delta_3^*) = \ln L_B(\delta_3^*) + \ln L_R(\delta_3^*) + \ln L_G(\delta_3^*). \quad (11)$$

Then we apply the criterion of the maximum log-likelihood function. We look for the maximum value among log-likelihood functions for the first, second, and third classes in 3 spectral bands:

$$\max\left\{\ln L_1(\delta_1^*), \ln L_2(\delta_2^*), \ln L_3(\delta_3^*)\right\}.$$

Depending on which value of the log-likelihood function for which class is maximized, the test sample will belong to the same class.

V. CONCLUSIONS

In this paper, we presented a detailed methodology for land cover classification of Sentinel-2 multispectral imagery using a maximum likelihood classifier based on Rayleigh distribution models of pixel intensities. By applying histograms of training pixels from three classes ("Water", "Vegetation", "Buildings") and fitting Rayleigh distributions, we derived probabilistic models that characterize each class's spectral behavior in the Red, Green and Blue bands. The results demonstrate that the Rayleigh-based model can effectively distinguish between the chosen classes in the test scene vegetation by its high green reflectance. The process of comparing log-likelihoods for each class was shown to provide a clear decision rule, and the example calculations illustrate why the chosen class is the most probable given the pixel's signature. We also provided insight into how the Rayleigh distribution's one-parameter nature captures the essence of skewed reflectance distributions. From a methodological perspective, this work highlights how classical statistical techniques can be applied in a remote sensing context with minimal resources – open-source tools like SNAP and software like Excel. The approach is scientifically transparent and rooted in well-established probability theory, making it a useful exercise for preliminary analysis. Future work could expand on this foundation by incorporating more spectral bands and additional classes, as well as by comparing the performance of the Rayleigh MLE classifier against more common approaches (Gaussian MLE, or even modern machine learning classifiers). It would be interesting to investigate hybrid models where some classes use Rayleigh distributions and others use Gaussians or other distributions as appropriate. In conclusion, the study provides a comprehensive example of maximum likelihood classification using an alternative statistical model for class-conditional densities. The successful classification of Sentinel-2 imagery into vegetation category using the Rayleigh distribution attests to the flexibility of the MLE framework. This work contributes to the broader theme of probabilistic remote sensing, illustrating how probabilistic models, when appropriately applied, can yield effective land cover mapping results.

REFERENCES

[1] I. J. Myung, “Tutorial on maximum likelihood Estimation,” *Journal of Mathematical Psychology*, 47 (1), pp. 90–100, 2003. https://doi.org/10.1016/S0022-2496(02)00028-7

[2] D. Lu and Q. Weng, “A survey of image classification methods and techniques for improving classification performance,” *International Journal of Remote Sensing,* 28(5), 823–870, 2007. https://doi.org/10.1080/01431160600746456

[3] N. Balakrishnan, “Approximate MLE of the scale parameter of the Rayleigh distribution with censoring,” *IEEE Transactions on Reliability*, 38, 355–357, 1989. https://doi.org/10.1109/24.44181

[4] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, & B. A. Johnson, “Deep learning in remote sensing applications: A meta-analysis and review,” *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177, 2019. https://doi.org/10.1016/j.isprsjprs.2019.04.015

[5] A. F. H. Goetz, G. Vane, J. E. Solomon, & B. N. Rock, “Imaging spectrometry for Earth remote sensing,” *Science*, 228(4704), 1147–1153, 1985. https://doi.org/10.1126/science.228.4704.1147

[6] C. Li, J. Wang, L. Wang, L. Hu, P. Gong, “Comparison of Classification Algorithms and Training Sample Sizes in Urban Land Classification with Landsat Thematic Mapper Imagery,” *Remote Sensing,* 6, 964–983, 2014. https://doi.org/10.3390/rs6020964

[7] C. Persello and L. Bruzzone, “Active and Semisupervised Learning for the Classification of Remote Sensing Images,” *IEEE Transactions on Geoscience and Remote Sensing,* 52, 6937–6956, 2014. https://doi.org/10.1109/TGRS.2014.2305805

[8] B. Janga, G. P. Asamani, Z. Sun, N. C. Cristea, “A Review of Practical AI for Remote Sensing in Earth Sciences,” *Remote Sensing*, 15, 2023 https://doi.org/10.3390/rs15164112

[9] I. Prokopenko, S. Alpert, Y. Petrova, “Applications of Dempster – Shafer evidence theory to data processing in remote sensing,” *CEUR Workshop Proceedings*, 65–76, 2025.

[10] C. Toth, and G. Jóźków, “Remote sensing platforms and sensors: A survey,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 115, pp. 22–36, May 2016, https://doi.org/10.1016/j.isprsjprs.2015.10.004

Received March 27, 2025

Prokopenko Igor. ORCID 0000-0003-4169-3774. Doctor of Engineering Science. Professor.
Department of Telecommunications and Radioelectronic Systems, Faculty of Aeronautics, Electronics and Telecommunications, State University “Kyiv Aviation Institute”, Kyiv, Ukraine.
Education: Kyiv Institute of Civil Aviation Engineers, Kyiv, Ukraine, (1972).
Research area: statistical processing of signals and data.
Publications: more than 250 scientific papers.
E-mail: igorprok48@gmail.com

Alpert Sofiia. ORCID 0000-0002-7284-6502. Candidate of Sciences (Engineering). Assistant Professor.
Department of Aerospace Geodesy and Land Management, Architecture, Construction and Design Faculty, State University “Kyiv Aviation Institute”, Kyiv, Ukraine
Education: National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine (2011).
Research area: remote sensing of the Earth, mathematical methods for processing aerospace images, solving tasks in the field of radar surveillance
Publications: more than 80 papers.
E-mail: sonyasonet87@gmail.com

Alpert Maksym. ORCID 0000-0002-8938-1473. PhD.
The Department of Information Systems and Technologies, Faculty of Informatics and Software Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.
Education: National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine, (2025).
Research area: Strategic priority area of innovation activity: Development of modern information and communication technologies, robotics.
Publications: 6 papers.
E-mail: max292009@gmail.com

Dmytruk Anastasiia. ORCID 0000-0002-6410-4680. Postgraduate student.
Department of Telecommunications and Radioelectronic Systems, Faculty of Aeronautics, Electronics and Telecommunications, State University “Kyiv Aviation Institute”, Kyiv, Ukraine
Education: National aviation universe, Kyiv, Ukraine.
Research area: statistical signal processing.
Publications: more than 10 scientific papers.
E-mail: gyhyre@gmail.com

І. Г. Прокопенко, С. І. Альперт, М. І. Альперт, А. Ю. Дмитрук. Класифікування знімку Sentinel-2 із використанням моделі релеєвського розподілу

На даний час класифікування земного покриву за супутниковими знімками є однією з найактуальніших та найважливіших задач дистанційного зондування. Багатоспектральні супутникові знімки, такі як знімки Sentinel-2, забезпечують зображення високої роздільної здатності у різних спектральних каналах, що дозволяє детально розрізняти наземні об’єкти. У даній роботі представлено метод класифікування супутникових знімків на основі розподілу Релея, методу максимальної правдоподібності та функцій правдоподібності. Розглянуто три класи земного покриву, а саме: “Вода”, “Рослинність” та “Забудови” та три спектральні канали (червоний, зелений та синій спектральний канал). Запропонована процедура класифікування включає моделювання розподілу спектру із використанням Релеєвського розподілу. Параметри розподілу Релея для кожного класу та для кожного спектрального каналу оцінюються на основі навчальних даних за наведеною формулою. Для обробки зображень використовується програмне забезпечення ESA SNAP. Для процедури класифікування використовується метод максимальної правдоподібності. У дистанційному зондуванні цей метод використовується для класифікування пікселів на супутникових знімках за різними класами. Цей метод базується на віднесенні кожного пікселя тому класу, до якого він належить із найбільшою ймовірністю. Описується методологія, яка включає в себе підготовку даних із використанням ESA SNAP та аналіз даних у Microsoft Excel. Розглянуто математичне формулювання розподілу Релея та алгоритм розрахунку функцій правдоподібності для кожного класу та для кожного спектрального каналу. Результати включають параметри розподілу Релея для кожного класу та для кожного спектрального канала, карти класифікації, розрахунок функцій правдоподібності та результат процедури класифікування. Результат класифікування залежить від того, якому саме класу відповідає максимальна функція правдоподібності. Було розглянуто приклад, де із використанням методу максимальної правдоподібності та розподілу Релея визначається клас “Рослинність”. Запропонований підхід може бути застосований для класифікування земного покриву, екологічного моніторингу, сільськогосподарських та геологічних задач.

Ключові слова: функції правдоподібності; метод максимальної правдоподібності; розподіл Релея; дистанційне зондування; класифікування супутникового зображення; обробка сигналів спектральних каналів супутникових зображень.

Прокопенко Ігор Григорович. ORCID 0000-0003-4169-3774. Доктор технічних наук. Професор.
Кафедра телекомунікаційних та радіоелектронних систем, Факультет аеронавігації, електроніки та телекомунікацій, Державний університет “Київський авіаційний інститут”, Київ, Україна.
Освіта: Київський інститут інженерів цивільної авіації, Київ, Україна, (1972).
Напрям наукової діяльності: статистична обробка сигналів і даних.
Кількість публікацій: більше 250 наукових робіт.
E-mail: igorprok48@gmail.com

Альперт Софія Іоганівна. ORCID 0000-0002-7284-6502. Кандидат технічних наук. Доцент.
Кафедра аерокосмічної геодезії та землеустрою, Факультет архітектури, будівництва та дизайну, Державний університет “Київський авіаційний інститут”, Київ, Україна.
Освіта: Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна, (2011).
Напрям наукової діяльності: дистанційне зондування Землі, математичні методи обробки аерокосмічних зображень, вирішення задач у сфері радіолокаційного спостереження
Кількість публікацій: більше 80 наукових робіт.
E-mail: sonyasonet87@gmail.com

Альперт Максим Іоганович. ORCID 0000-0002-8938-1473. Доктор філософії.
Кафедра інформаційних систем та технологій, Факультет інформатики та обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.
Освіта: Національний технічний університет України “Київський політехнічний інститут імені Ігоря Сікорського”, Київ, Україна, (2025).
Напрям наукової діяльності: Стратегічний пріоритетний напрям інноваційної діяльності: Розвиток сучасних інформаційних, комунікаційних технологій, робототехніки
Кількість публікацій: 6 наукових робіт.
E-mail: max292009@gmail.com

Дмитрук Анастасія Юріївна. ORCID 0000-0002-6410-4680. Аспірантка.
Кафедра телекомунікаційних та радіоелектронних систем, Факультет аеронавігації, електроніки та телекомунікацій, Державний університет “Київський авіаційний інститут”, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна.
Напрям наукової діяльності: статистична обробка сигналів.
Кількість публікацій: більше 10 наукових робіт.
E-mail: gyhyre@gmail.com