UDC 004.855.5(045)
DOI:10.18372/1990-5548.76.17662

[1]**V. O. Nikitin,**
[2]**V. Y. Danilov**

INTEGRATION OF FRACTAL DIMENSION IN VISION TRANSFORMER FOR SKIN CANCER CLASSIFICATION

[1,2]Educational and scientific complex "Institute for Applied System Analysis", National Technical University of Ukraine “Ihor Sikorsky Kyiv Polytechnic Institute,” Kyiv, Ukraine
E-mails: [1]nvo63911@gmail.com ORCID 0009-0001-9921-0213,
[2]danilov1950@ukr.net ORCID 0000-0003-3389-3661

Abstract**—***In order to classify skin lesions, many efforts have been made to create various automated diagnostic systems. For that purpose many efforts have been put in creating various automated diagnostics systems Nowadays, with the rapid advancements in deep learning, Vision Transformers have emerged as powerful models for image processing and analysis purposes. This type of model has already proved useful for cancer detection and classification tasks in particular. However, the complexity and variability of skin lesions present significant challenges in accurately classifying them. Integrating the concept of fractal dimension into Vision Transformers can potentially improve their performance by capturing the intricate structural patterns of skin lesions. This paper aims to explore the integration of fractal dimension metrics into a Vision Transformer for skin cancer classification. The problem at hand is to investigate the integration of fractal dimension metrics into the existing Vision Transformer architecture for the accurate classification of skin lesions as cancerous or non-cancerous. Fractal dimensions provide a measure of the complexity and irregularity of an object, which can be informative in characterizing skin cancer lesions. We aim to research possability and ways of incorporating fractal dimension metrics into the Vision Transformer model for results improvements.*

Index Terms—Machine learning; skin cancer; skin lesion classification; Vision Transformer; fractal dimension; classification tasks.

I. INTRODUCTION

Skin cancer is a prevalent and widely diagnosed disease worldwide, with its incidence increasing at an alarming rate. However, it is important to note that the true number of cases remains unknown, as not all instances are reported to the dedicated departments responsible for collecting statistical data. The significance of early detection in improving the chances of a full recovery cannot be overstated. Therefore, there is a pressing need for robust and accurate classification systems that aid in the early detection and diagnosis of skin lesions associated with cancer [1], [2].

In recent years, significant advancements in deep learning techniques have paved the way for new and promising models for image processing and analysis tasks. Among these, the Vision Transformer (ViT) has emerged as a highly potent model, showcasing its effectiveness in various domains, including cancer detection and classification. In fact, prior research by one of the authors has already explored the application of ViT in skin cancer classification, demonstrating its potential [3].

While ViT has demonstrated superiority over traditional approaches such as convolutional neural networks (CNNs), accurately classifying skin lesions with high complexity and variability remains a challenging task. To address this, we propose the integration of an additional metric, specifically the concept of fractal dimension, into the ViT architecture. Fractal dimension serves as a valuable measure of complexity and irregularity in objects, offering the potential to capture the intricate structural patterns inherent in skin lesions.

Skin lesions exhibit fractal properties, and their self-similarity can be quantified using fractal dimension. By incorporating fractal dimension metrics into the ViT model, we aim to enhance its performance in skin cancer classification tasks. This research aims to investigate three different approaches for integrating the fractal dimension of skin lesions into the Vision Transformer.

Through this study, we seek to contribute to the advancement of automated diagnostics systems for dermatological diseases, specifically in the context of skin cancer classification. By exploring the integration of fractal dimension metrics into the ViT model, we aim to pave the way for more accurate and reliable early detection systems, ultimately leading to improved patient outcomes and a greater chance of successful treatment.

II. Vision Transformer

Vision Transformer (ViT) is a deep learning architecture that has gained significant attention in the field of computer vision that was first introduced in the paper “An Image is Worth 16x16 words: Transformers for Image Recognition at Scale” [3]. It represents a departure from traditional Convolutional Neural Networks (CNNs) by utilizing a transformer-based model, originally introduced for natural language processing (NLP) tasks, and applying it to image data. ViT has shown impressive performance on various visual recognition tasks, including image classification, object detection, and segmentation. In this description, we will delve into the workings of ViT, its key components, and its advantages over traditional CNN-based models. The transformer architecture, initially proposed for sequence tasks like machine translation, consists of an encoder-decoder framework with self-attention mechanisms. To apply the transformer architecture to images, ViT divides an input image into a sequence of fixed-size patches, treating them as tokens similar to words in NLP. Each patch is then linearly projected into a lower-dimensional representation known as embeddings. These embeddings, along with positional encodings, are fed into the transformer encoder [3].

The core idea behind transformers is self-attention, a mechanism that allows the model to weigh the importance of different positions within the input sequence. This attention mechanism enables the model to focus on relevant parts of the input and effectively capture global dependencies. Self-attention is a mechanism that computes attention weights for each position in the input sequence based on its relationships with all other positions. Multi-head attention, a variation of self-attention, performs multiple attention operations in parallel. This allows the model to capture different types of relationships and learn more diverse representations [4].

The final hidden states of the transformer encoder are used for classification. A simple linear layer is appended on top of the encoder to map the learned representations to class probabilities. During training, the model is optimized to minimize a suitable loss function, such as cross-entropy, based on the predicted probabilities and ground-truth labels.

III. Fractal Dimension

Fractals are complex mathematical objects that exhibit self-similarity across different scales. Fractal dimension is a measure that quantifies the level of complexity and intricacy within a fractal. Fractals can be found in various natural and man-made phenomena, such as coastlines, clouds, trees, and even financial markets. They possess unique properties that make them intriguing for studying and understanding complex systems. The idea of fractal dimension emerged from the need to measure and characterize the intricate structure of fractals. Unlike traditional geometric objects, which have integer dimensions (e.g., a line has dimension 1, a plane has dimension 2), fractals exhibit non-integer dimensions [5].

The concept of fractal dimension can be explained by considering the relationship between the scale at which we observe a fractal and the level of detail we can perceive. As we zoom in on a fractal, we discover smaller copies of the overall pattern, repeating itself in a self-similar manner. The fractal dimension quantifies the extent of this self-similarity across different scales [6].

There are several methods to calculate the fractal dimension of a given object or dataset. One commonly used method is the box-counting technique. In this approach, we cover the fractal with a grid of equally sized boxes and count the number of boxes that intersect with the fractal. By varying the size of the boxes, we can observe how the number of intersecting boxes changes. The fractal dimension is then determined by analyzing the scaling relationship between the box size and the number of intersecting boxes.

In this article we try to use a box-counting method for calculating fractal dimension on skin lesions dataset and use this metric for an improvement of classification results we’re getting using specific architecture of a ViT.

IV. Dataset

Dataset HAM10000 was used for training the model. The dataset, available on Harvard Dataverse, comprises 10000 skin lesions [7]. To ensure balanced training and evaluation, we implemented a data splitting strategy based on the "cancer – not cancer" criteria. Recognizing the significance of addressing imbalanced classes, we carefully structured the training set by selecting a proportionate number of samples from each group. This approach allowed the model to learn from a diverse range of cases while maintaining equal representation of both classes.

The concrete steps taken for images preprocessing are resize to 100x100 and normalization. For fractal dimension calculation images we converted the images to grayscale, applied a Gaussian blur to the grayscale image using a kernel size of (3, 3). This blurring operation helped to reduce noise and smooth out irregularities in the image, resulting in a cleaner representation. After that we applied canny edge detection – an image processing technique that identifies and extracts the edges of objects within an image based on intensity gradients [8] (Fig. 1). The lower and upper thresholds were determined based on the calculated minimum, mean, and maximum values of the grayscale image.

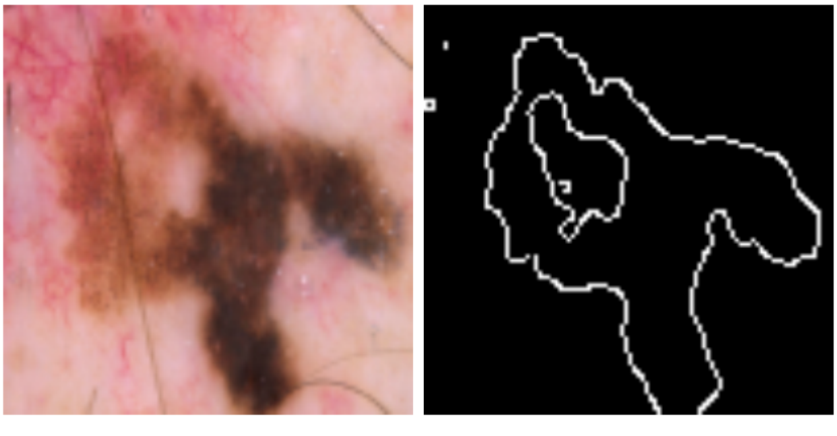

Fig. 1. Image from HAM1000 [5] dataset: (a) before and (b) after preprocessing for fractal dimension calculation

V. Calculation of the Fractal Dimension

The box counting method also known as Minkowski–Bouligand dimension is a technique used to estimate the fractal dimension of an object by measuring the number of boxes required to cover the object at different scales or resolutions. For instance, when applied to the British coastline, the box counting method involves progressively dividing the coastline into smaller boxes and counting the number of boxes needed to cover it at each scale, revealing the self-similar nature of the coastline and providing an estimation of its fractal dimension, which characterizes its intricate and complex structure at different levels of magnification [5].

It was chosen as the one that takes R^2 input by default, quick to implement and shows significant quality of the results (Fig. 2).

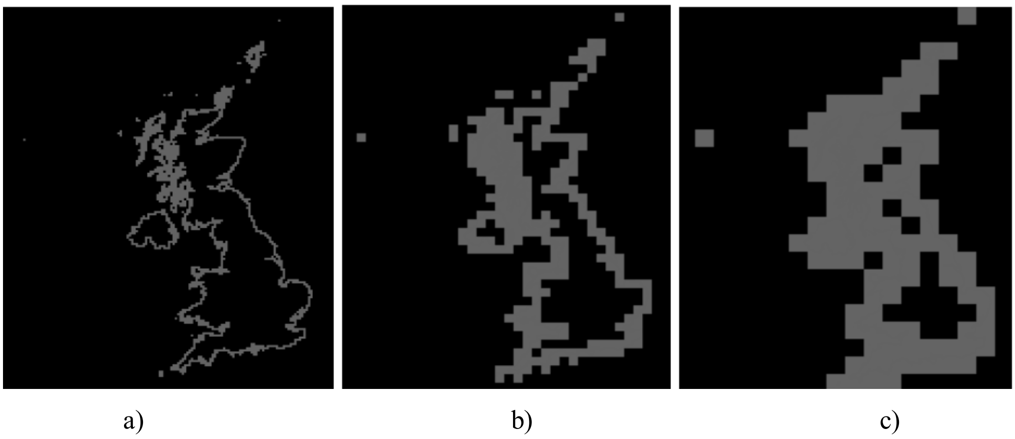

Fig. 2. Application of Minkowski–Bouligand dimension on Britain coastline with box sizes (a) – 2; (b) – 4; (c) – 8

VI. Used Model Architectures

Skin cancer classification using the ViT architecture has shown promising results, as demonstrated in a previous publication titled "Vision Transformer for Skin Cancer Classification" [9]. This model utilizes an improved attention function, which enhances the detection of skin lesion edges. However, further improvements are required to enhance the performance of the ViT model by incorporating the fractal dimension metric.

In this work, we aim to introduce enhancements to the existing ViT architecture by integrating the fractal dimension metric. The fractal dimension provides valuable information about the intricate structures present in skin lesions, which can aid in accurate classification. To effectively compare the performance of these architectures, we also trained a model without the integration of fractal dimension, referred to as CancerViT.

We considered and tested 3 options of adding fractal dimension into the model. First is to add it as a patch class at the stage of patch linearization. This model will be referenced as PatchFDViT. That way we treat fractal dimension as part of the original image. As for the second architecture, we've added a fractal dimension in the transformer itself, in every layer. It's going to be called AttentionFDViT. It should specifically emphasize the weight of this metric for the attention function. In the third and final model we've added fractal dimension as additional input for the classification layer itself. The name of this model will be ClassFDViT. Therefore it's treated as a completely separate metric (Figs 3 and 4).

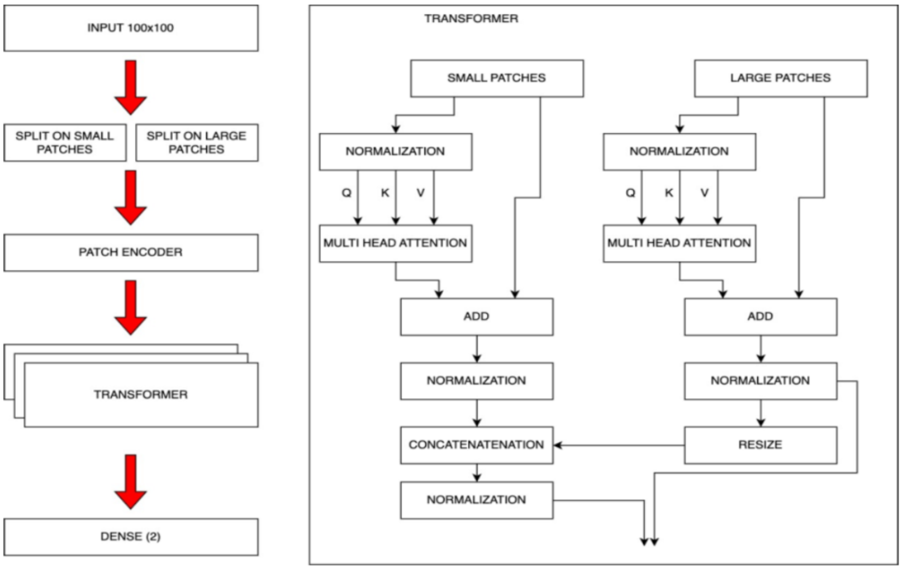

Fig. 3. Architecture of ViT model specified for skin cancer classification [9]

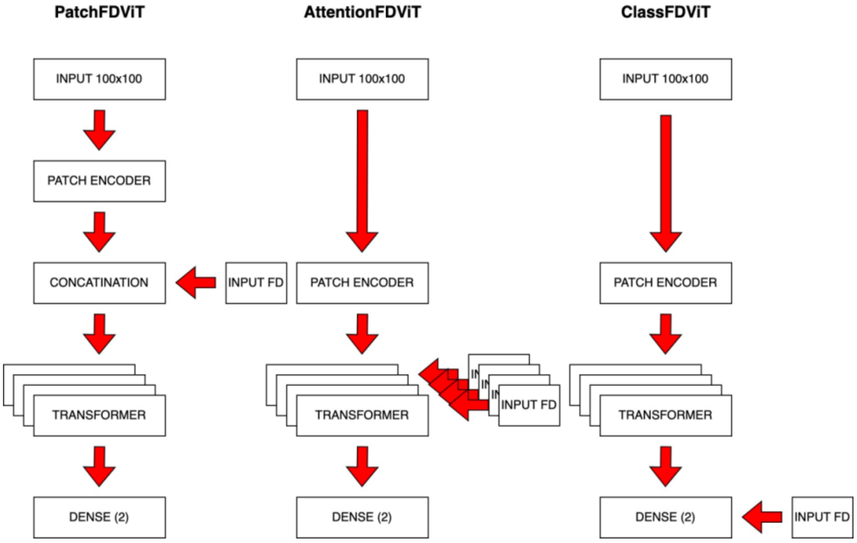

Fig. 4. Three considered alternatives of integrating the fractal dimension into the mode

VII. Results

In evaluating the performance of our classification models, we employ two metrics for comparison. The first metric is accuracy, which serves as a general measure of the model's overall classification quality. However, considering the medical nature of the task at hand, our primary focus lies in minimizing the occurrence of type 2 errors, as the cost of misclassifying cancerous lesions for patients is considerably higher than the reverse scenario.

To address this concern, we introduce the False Negative Rate (FNR) as our second metric. The

FNR quantifies the proportion of actual positive instances (cancerous lesions) that are incorrectly classified as negative (non-cancerous). It is computed by dividing the number of false negatives by the sum of false negatives and true positives.

By incorporating the FNR metric into our evaluation, we aim to assess the model's performance specifically in reducing the misclassification of cancerous lesions. This approach ensures that our analysis accounts for the critical aspect of minimizing false negatives, which directly impacts patient outcomes and prevents potential delays in diagnosis and treatment.

As can be seen from Table I only PatchFDViT showed worsening of the both metrics compared to CancerViT. ClassFDViT outperformed CancerViT by FNR and AttentionFDViT showed both metrics increased.

TABLE I RESULTS OF THE LEARNING

Model Name	Accuracy, %	FNR, 0-1
CancerViT	78.14	0.32
PatchFDViT	76.05	0.39
AttentionFDViT	79.04	0.23
ClassFDViT	78.14	0.27

VIII. CONCLUSIONS

Among the three considered architectures, two of them exhibited noteworthy improvements in performance, while the remaining architecture, along with the control model, did not show any substantial enhancement. Notably, the architectures that integrated the fractal dimension metric demonstrated noticeable improvements in the evaluation metrics.

These findings indicate that the inclusion of the fractal dimension of skin lesions within the attention layers of the model plays a crucial role in improving the quality of skin cancer classification when using vision transformers. By leveraging the informative nature of fractal dimension measurements, the model becomes more adept at capturing and interpreting the intricate structural patterns inherent in skin lesions. Consequently, this integration enhances the model's ability to accurately classify skin lesions as cancerous or non-cancerous.

IX. FUTURE INVESTIGATIONS

This work opens up several promising directions for future research and enhancement. Here are a few examples:

Firstly, it is essential to explore different approaches for measuring the fractal dimension of skin lesions, including machine learning-based methods. This is crucial because the precise determination of lesion boundaries can sometimes be subjective, leading to variations in results. By employing alternative preprocessing and measuring techniques, we can potentially obtain different fractal dimension measurements, improving metrics.

Secondly, in this study, a specific ViT architecture with a predetermined set of parameters, known for its effectiveness in skin cancer classification, was utilized. However, for further advancements, it is necessary to consider other variations of ViT architectures with different parameter settings. It is plausible that an architecture specifically tailored to integrate the fractal dimension metric may yield even higher performance. For example, CrossViT [10] is similar to the model used in this paper and may also be effective and perform the task.

In summary, the research topic of integrating fractal dimension into Vision Transformers for skin cancer classification presents numerous avenues for further exploration and improvements. Investigating alternative measurement approaches and exploring a broader range of ViT architectures with different parameter configurations can potentially uncover novel insights and lead to enhanced performance in accurately classifying skin lesions. There is ample scope for future discoveries and advancements in this field.

REFERENCES

[1] *Cancer facts & figures 2023* (no date) *American Cancer Society*. Available at: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2023-cancer-facts-figures.html (Accessed: 10 June 2023).

[2] (No date) *Adjusted rates 2018 melanoma of skin C43 table 1 – general rates, 2018*. Available at: http://www.ncru.inf.ua/publications/BULL_21/PDF_E/38-39-mel.pdf (Accessed: 10 June 2023).

[3] A. Dosovitskiy, *et al. An image is worth 16x16 words: Transformers for image recognition at scale*, (2021), *arXiv.org*. Available at: https://doi.org/10.48550/arXiv.2010.11929 (Accessed: 10 June 2023).

[4] A. Vaswani, *et al. Attention is all you need*, (2017), *arXiv.org*. Available at: https://doi.org/10.48550/arXiv.1706.03762 (Accessed: 10 June 2023).

[5] K. J. Falconer, (2003). *Fractal geometry: Mathematical foundations and applications: K. J. Falconer, onlinelibrary.wiley.com*. Available at: https://doi.org/10.1002/0470013850 (Accessed: 10 June 2023).

[6] M. Kirkby, (1983). The fractal geometry of nature. Benoit B. Mandelbrot. W. H. Freeman and co., San Francisco, 1982. No. of pages: 460. Price: £22.75 (hardback). Earth Surface Processes and Landforms,

8(4), 406. https://doi.org/10.1002/esp.329008041 Accessed: 10 June 2023).

[7] P. Tschandl, (2023). *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions*, *Harvard Dataverse*. Available at: https://doi.org/10.7910/DVN/DBW86T (Accessed: 10 June 2023).

[8] J. F. Canny, (1986). *A computational approach to edge detection – researchgate*. Available at: http://dx.doi.org/10.1109/TPAMI.1986.4767851 (Accessed: 10 June 2023).

[9] V. Nikitin, and N. Shapoval, (2023). *Vision Transformer for skin cancer classification*, *Scientific Collection 'InterConf+'*. Available at: https://doi.org/10.51582/interconf.19-20.05.2023.039 (Accessed: 10 June 2023).

[10] C.-F. Chen, Q. Fan, and R. Panda, (2021). *Crossvit: Cross-attention multi-scale vision transformer for Image Classification*, arXiv.org. Available at: https://arxiv.org/abs/2103.14899 (Accessed: 10 June 2023)

Received March 11, 2023

Nikitin Vladyslav. MSc in Computer Science. ORCID 0009-0001-9921-0213.
Educational and scientific complex "Institute for Applied System Analysis", National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.
Education: National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine (2022).
Research area: deep learning, computer vision, transformers.
Publications: 1.
E-mail: nvo63911@gmail.com

Danilov Valery. Doctor of Engineering Science. Professor. ORCID 0000-0003-3389-3661.
Educational and scientific complex "Institute for Applied System Analysis", National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
Education: Kyiv Order of Lenin Polytechnic Institute named after the 50th anniversary of the Great October Socialist Revolution, Ukraine, (1972).
Research area: theory of filtration and control of systems with distributed parameters, methods and algorithms for optimizing the processing of hydroacoustic signals, synthesis of hydroacoustic antennas, computer vision.
Publications: 78.
E-mail: danilov1950@ukr.net

В. О. Нікітін, В. Я. Данилов. Інтеграція фрактальної розмірності у візуальний трансформер для класифікації раку шкіри

З метою класифікації уражень шкіри було зроблено багато зусиль для створення різноманітних автоматизованих систем діагностики. На сьогоднішній день, зі стрімкими досягненнями у глибокому навчанні, візуальні трансформери виходять на передній план як потужні моделі для обробки та аналізу зображень. Цей тип моделей вже довів свою корисність для виявлення та класифікації ракових захворювань зокрема. Однак, складність і змінність уражень шкіри створюють значні виклики при точній класифікації. Інтеграція концепції фрактальної розмірності у візуальні трансформери може покращити їхню продуктивність, захоплюючи складні структурні зразки уражень шкіри. Метою цієї роботи є дослідження інтеграції метрик фрактальної розмірності у візуальний трансформер для класифікації раку шкіри. Проблема, яку необхідно дослідити, полягає у вивченні можливості і способів інтеграції метрик фрактальної розмірності у існуючу архітектуру візуального трансформера для точної класифікації уражень шкіри як ракових або нераковых. Фрактальні розмірності надають міру складності та неправильності об'єкта, що може бути інформативним при характеризації уражень шкіри, пов'язаних з раком. Планується дослідити можливості та шлях.

Ключові слова: машинне навчання; рак шкіри; класифікація пухлин шкіри; візуальний трансформер; фрактальна розмірність; задачі класифікації.

Нікітін Владислав Олегович. Магістр комп'ютерних наук. ORCID 0009-0001-9921-0213.
Навчально-науковий комплекс «Інститут прикладного системного аналізу», 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. (2022).
Напрям наукової діяльності: глибоке навчання, комп'ютерний зір, трансформери.
Кількість публікацій: 1.
E-mail: nvo63911@gmail.com

Данилов Валерій Якович. Доктор технічних наук. Професор. ORCID 0000-0003-3389-3661.
Навчально-науковий комплекс «Інститут прикладного системного аналізу», Національний технічний університет України «Київський Політехнічний Інститут імені Ігоря Сікорського», Київ, Україна.
Освіта: Київський ордена Леніна політехнічний інститут імені 50-річчя Великої Жовтневої соціалістичної революції, Україна, (1972).
Напрям наукової діяльності: теорія фільтрації та керування системами з розподіленими параметрами, методи й алгоритми оптимізації обробки гідроакустичних сигналів, синтезу гідроакустичних антен, комп'ютений зір.
Кількість публікацій: 78.
E-mail: danilov1950@ukr.net