

D. Hlavcheva<sup>1</sup>, V. Yaloveha<sup>1</sup>, A. Podorozhniak<sup>1</sup>, N. Lukova-Chuiko<sup>2</sup>

<sup>1</sup>National Technical University “Kharkiv Polytechnic Institute”, Kharkiv, Ukraine

<sup>2</sup>Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

## A COMPARISON OF CLASSIFIERS APPLIED TO THE PROBLEM OF BIOPSY IMAGES ANALYSIS

**Abstract.** The purpose of the research is to compare classification algorithms for the histopathological images analyzing issue and to optimize the parameters for obtaining better classification accuracy. The following tasks are solved in the article: preprocessing of BreCaHAD dataset images, implementation and training of CNN, applying K-nearest neighbours, SVM, Random Forest, XGBoost, and perceptron algorithms for classifying features that were extracted by CNN, and results comparison. The object of the research is the process of classifying tumor cells in the microscopic biopsy images. The subject of the research is the process of using ML algorithms for classification of the features extracted by CNN from input biopsy image. The scientific novelty of the research is a comparative analysis of classifiers on the task of “tumor” and “healthy” cells images classification from processed BreCaHAD dataset. As a result it was obtained that from chosen classifiers SVM reached the highest accuracy on test data – 0.972. This is the only algorithm that shows better accuracy than perceptron. Perceptron gets 0.966 classification accuracy. K-nearest neighbours, Random Forest, and XGBoost algorithms reached lower results. The algorithms' hyperparameters optimization was carried out. The results have been compared with related works. The following research methods are used: the theory of deep learning, mathematical statistics, parameters optimization.

**Keywords:** machine learning; deep learning; convolutional neural network; grid search; biopsy.

### Introduction

Machine learning (ML) is defined as a set of methods that automatically detect patterns in data, and then utilize the uncovered patterns to predict future data or enable decision making under uncertain conditions. The most representative characteristic of ML is that it is driven by data, and the decision process is accomplished with minimum interventions by a human [1].

Machine learning technology powers many aspects of modern society including different medicine tasks such as medical images classification, segmentation, and object detection. Also, different ML approaches are used for analyzing patients' data to define risk groups for diseases, choose better treatment strategies, etc.

Many diagnostic tasks require an initial search process to detect abnormalities, and to quantify measurements and changes over time. Computerized tools, specifically image analysis and machine learning, are the key enablers to improve diagnosis, by facilitating identification of the findings that require treatment and to support the expert's workflow. Among these tools, deep learning is rapidly proving to be the state-of-the-art foundation, leading to improved accuracy [2].

Image or exam classification was one of the first areas in which deep learning made a major contribution to medical image analysis [3].

Other classifiers were subsequently developed, including decision trees, boosting and support vector machines. Each of these has been applied to medical image analysis [2].

According to “Cancer Tomorrow” [4] number of cancer cases in the world will increase from 18.1 million in 2018 to 29.5 million in 2040. In Ukraine the number of new cases in 2018 is almost 170 thousand for both sexes and all ages. The number of death is almost 100 thousand [5].

So, one of the important medical areas in which ML approaches is being used is cancer research.

A variety of ML techniques, including Artificial Neural Networks (ANNs), K-nearest neighbours, Support Vector Machines (SVMs) and Decision Trees (DTs) have been widely applied in cancer research for the development of predictive models, resulting in effective and accurate decision making [6].

The clinical management of cancer is informed by histopathological evaluation of biopsy tissues, wherein thin sections of a biopsy are processed to visualize tissue and cell morphologies for signs of disease.

A number of image analysis tasks in digital pathology involve the quantification and highlight of morphological features (e.g., cell or mitotic count, nuclei grading, epithelial glands morphology). These tasks invariably require the identification of histologic primitives (e.g., cell, nuclei, mitosis, epithelium, cellular membranes, etc.) [7]. After problem review the main purpose of the research was defined. It is the comparison of different ML algorithms and perceptron for classifying tumor and healthy cells in microscopic biopsy images.

According to the purpose the research tasks are:

- preparation of microscopic biopsy images dataset;
- implementation and training of CNN;
- applying of chosen algorithms for classifying features extracted by CNN;
- comparison of obtained results.

In the paper, the Convolutional Neural Network has been used to extract features for classifiers. The classifier has been compared. Such an approach can improve the accuracy of tumor and healthy cells classification for cancer stage diagnosis.

### 1. Classification algorithms

#### 1.1 K-nearest Neighbours

K-nearest Neighbour (*k*-NN) algorithm classifies examples based on the class of their nearest neighbours.

There are several ways in which the *k* nearest neighbours can be used to determine the class. The

simple approach is to define a class among the nearest neighbours and assign more weight to the nearer neighbours in deciding the class. This can be achieved by distance weighted voting where the neighbours get to vote on the class [8]. The  $k$  nearest neighbours are selected based on this distance metric. Usually Minkowski Distance metric is used as basic. The formula for the Minkowski distance is:

$$\rho(y, x_i) = \left( \sum_j |y_j - x_{ij}|^p \right)^{1/p}. \quad (1)$$

### 1.2 Support Vector Machine

Last years the algorithm became popular and it is used for solving regression and classification problems [9, 10]. The main idea is divorced to find a decision surface where the margin of separation between classes is maximized or to minimize a quadratic function with certain constraints:

$$\begin{aligned} \min_{\omega, b, \varepsilon} \quad & \left( \|\omega\|^2 / 2 + C \sum_j \varepsilon_j \right); \\ y_j (\omega^T \varphi(x_j) + b) \geq 1 - \varepsilon_j, \quad & \varepsilon_j \geq 0, \end{aligned} \quad (2)$$

where  $\omega$  - parameters of function,  $C$  - parameter that regularize penalty  $\varepsilon$  for point that are laying in wrong side of the margin boundary [11-13].

In the study several kernel functions are used such as linear, polynomial, sigmoid and radial-bias functions. This constrained optimization problem is solved by using Lagrange multipliers [13].

### 1.3 Random Forest

Decision Trees are a popular supervised learning algorithm used for classification issues. The main idea is to construct a model that can predict a target class by learning decision rules formed from the data features [14]. In simple case a decision tree is a binary structure used to classify a given input. A node in the tree represents a comparison against some field. Advantages of using decision trees include non-linearity, support for categorical variables, interpretability and robustness [15]. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [16]. In simple case Random Forest uses averaging to improve the predictive accuracy and control over-fitting [17].

### 1.4 XGBoost

Boosting is based on weighting the examples according to how hard they are to get right, and reward classifiers based on the weight of the examples they get right, not just the count. To set the weights of the classifier, it is needed to adjust the weights of the training examples.

Easy training examples will be properly classified by most classifiers because they get bigger reward for getting the hard cases right [15].

XGBoost is an ensemble tree method (like Gradient Boosting) and the method apply the principle of boosting for weak learners. However, XGBoost was introduced for better speed and performance. In-built cross-validation ability, efficient handling of missing data, regularization for avoiding overfitting, catch

awareness are common advantages of XGBoost algorithm [16].

## 2. Convolutional Neural Network

Deep learning is a specific subfield of machine learning: from data that puts an emphasis on learning successive layers of increasingly meaningful representations [18].

CNNs became one of the most popular neural networks last years. They are widely used for image analyzing [3] and classification issues [19, 20].

A typical CNN is composed of an input layer and an output layer as well as many hidden layers that are alternately stacked convolution, normalization, spatial pooling and fully connected (i.e., dense) layers [21].

Convolutional layers consist of feature maps that encode different high- or low-level features using weights. Although the role of the convolutional layer is to detect local conjunctions of features from the previous layer, the role of the pooling layer is to merge semantically similar features into one. Pooling layer computes the maximum of one or few feature maps. Backpropagating gradients through a ConvNet is as simple as through a regular deep network, allowing all the weights in all the filter banks to be trained [22].

Deep CNNs automatically learn abstractions obtained from input data (in our case images). Recent results indicate that the generic descriptors extracted from CNNs are extremely effective in object recognition. Medical image analysis groups across the world are quickly entering the field and applying CNNs and other deep learning methodologies to a wide variety of applications [2].

So, in the research the CNN has been trained to get necessary features from the dataset automatically.

## 3. Results and Discussions

For the experiment implementation Python has been used as programming language. Because it supports a lot of additional modules that provides an opportunity to prepare dataset, implement neural network, use ML algorithms, optimize parameters, and measure results. Modules used in this research are listed below:

- PIL image – for slicing and preprocessing images from primary dataset;
- Numpy – for working with multi-dimensional arrays;
- Scikit learn – for implementation of different ML algorithms (including fitting and results evaluating);
- Tensorflow and Keras – for CNN creation, training and results evaluating.

CNN training and ML algorithms optimization experiments were launched in the Google Colabotarty [23].

In the Research the BreCaHAD dataset has been used. It is publicly available to the biomedical imaging community. The BreCaHAD dataset contains microscopic biopsy images which are saved in uncompressed (.TIFF) image format, three-channel RGB with 8-bit depth in each channel, and the dimension is 1360×1024 pixels and each image is annotated. The

annotations for the BreCaHAD dataset are provided in JSON (JavaScript Object Notation) format. In the given Table 1, Data file 4, the JSON file (ground truth) contains two mitosis and only one tumor nuclei annotations. Here,  $x$  and  $y$  are the coordinates of the centroid of the annotated object, and the values are between  $[0, 1]$  (divided by width and height of an image) [24].

In the study, the task of binary classification was posed. The classification was to determine whether the image contains a tumor nuclei or not. For the experiments primary dataset images have been sliced to images with size  $48 \times 48$  pixels. And after that, dataset consisted of

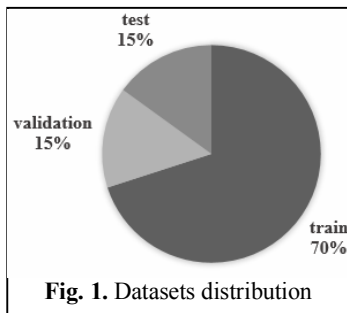


Fig. 1. Datasets distribution

40000 images (50% - with tumor nuclei and 50% - without). It was divided into train, validation and test as shown at Fig. 1.

For classifying images from dataset, classification system has been used. Its architecture is shown at Fig. 2. Designed system has 2 stages. The first stages

purpose is feature extracting from input image. It consists of 3 pairs of Convolutional and MaxPooling layers. These layers have 149,729 trainable parameters. In the end of this stage 128 features are highlighted from primary image. The second stage is a classifier that produces output result. Training of such classification system requires training of CNN with perceptron as classifier.

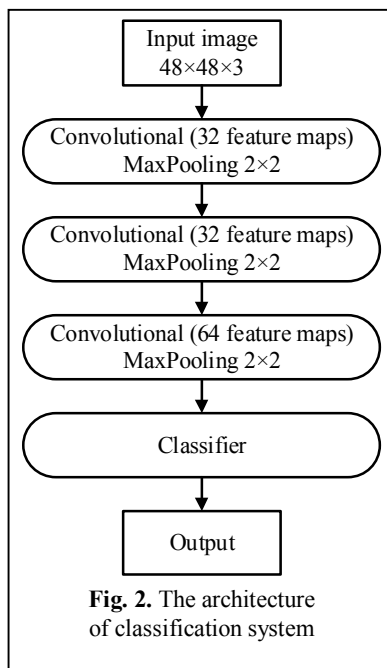


Fig. 2. The architecture of classification system

Then Convolutional and Max Pooling layers separated (the perceptron has been erased) and their outputs are used as input features for the ML algorithms. In the research these ML algorithms have been chosen: *K*-nearest Neighbours; Support Vector Machine; Random Forest; XGBoost. In the first stage of training it was obtained that perceptron as classifier showed 0.966 accuracy on test data.

Multi-dimensional grid search method has been used for the optimization of ML algorithms hyperparameters. In the  $m$ -dimensional parameter case, when we are given the initial region

$$I = [a_1; a_2] \times [b_1; b_2] \times \dots \in R^m,$$

we can consider the equi-distance grid [25]:

$$L = \left\{ (x_1, \dots, x_m)^T \mid \begin{aligned} x_1 &= a_1 + \frac{k_a}{n+1} (a_2 - a_1), k_a = 1, \dots, n, \\ x_2 &= b_1 + \frac{k_b}{n+1} (b_2 - b_1), k_b = 1, \dots, n, \dots \end{aligned} \right\}, \quad (3)$$

where  $n$  – number of experiments.

For each of the researched algorithms, the set of tuned parameters is different because of their organization principles. Final results of ML classification algorithms' accuracy score on test data are shown at the Table 1.

Table 1 – Accuracy on test data and best parameters for the ML algorithms

Algorithm	Parameter	Value	Accuracy
K-nearest neighbours	N neighbours	3	<b>0.963</b>
	Weight	uniform	
	Algorithm	auto	
	P	1	
SVM	Kernel	linear	<b>0.972</b>
	C	1.39e-5	
	Gamma	0.1	
Random Forest	Bootstrap	True	<b>0.953</b>
	Max depth	10	
	Min samples leaf	1	
	Min samples split	2	
	N estimators	70	
XGBoost	Min child weight	5	<b>0.954</b>
	Max depth	10	
	Gamma	0.3	
	Subsample	0.8	
	Colsample by tree	0.75	
	Reg alpha	0	
	N estimators	10	

Also, in this table all parameters and their values are presented and named in accordance with the documentation of Scikit learn python module [26].

In assume, SVM showed higher results than perceptron. Approach of using SVM as classifier of features extracted by CNN had been used in [27] and also showed higher results than perceptrone. SVM is an improvement of *k*-nearest neighbors approach and as shown in the Table 1 *k*-nearest neighbors algorithm reached less accuracy value than SVM.

Trees based algorithms (Random Forest and XGBoost) showed lower results than the perceptron. Also, these methods are more time-consuming than the others used in this research.

According to the previous authors' work [28] resulted classification accuracy on test data has been increased from 0.935 to 0.972. It can be explained by these reasons:

1. Increasing dimension of input images;
2. More complicated structure of CNN;
3. Using SVM as classifier;
4. Tuning of classifier parameters using grid search algorithm.

In the [29] comprehensive evaluation of 7 commonly used algorithms over 65 publicly available

gene expression datasets was held. As the result SVM and Random Forest presented better performances than the others algorithms researched in the paper.

Also, SVM and Artificial Neural Network reached the best accuracy in the research [30]. In the paper the ability of a multitude of linear and non-linear classification algorithms to discriminate between the electroencephalograms (EEGs) of patients with varying degree of Alzheimer's disease and their age-matched control subjects was explored. SVM showed 0.89 resulted accuracy in this task.

### Conclusions

Machine Learning, and especially Deep Learning approaches, are widely used in medical tasks. Automatization of medical images processing can possibly decrease time needed for image analyzing.

Cancer is a second leading cause of death worldwide [31]. Histopathological evaluation of biopsy images using ML approach can improve cancer taje diagnosis.

Based on the results of previous studies, the article proposes the comparative analysis of the K-nearest neighbours, SVM, Random Forest XGBoost and perceptron algorithms for classifying features extracted from the input image by Convolutional and MaxPooling layers.

It was obtained that SVM classifier performs the best accuracy – 0.972. SVM reached higher results than perceptron. Gained results have been compared with related works.

In further research multi classification will be performed on BreCaHAD datasets to classify all histological primitives on biopsy images.

### REFERENCES

1. Lee, J.-G., Jun, S., Cho, Y.-W., Lee, H., Kim, G. B., Seo, J. B. and Kim, N. (2017), "Deep learning in medical imaging: general overview", *Korean journal of radiology*, 18(4), pp. 570-584.
2. Greenspan, H., Van Ginneken, B. and Summers, R. M. (2016), "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique", *IEEE Transactions on Medical Imaging*, 35(5), pp. 1153-1159.
3. Laak, J. A. (2017), "A survey on deep learning in medical image analysis", *Medical image analysis*, 42, pp. 60-88.
4. *Cancer tomorrow* (2020), available at: <https://gco.iarc.fr/tomorrow/home>
5. *International Agency for Research of Cancer, World Health Organization Europe: Ukraine – Global Cancer Observatory* (2018), available at: <https://gco.iarc.fr/today/data/factsheets/populations/804-ukraine-fact-sheets.pdf>
6. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. and Fotiadis, D. I. (2015), "Machine learning applications in cancer prognosis and prediction", *Computational and structural biotechnology journal*, 13, pp. 8-17.
7. Janowczyk, A. and Madabhushi, A. (2016), "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases", *Journal of pathology informatics*, 7.
8. Cunningham, P. and Delany, S. J. (2020), *k-Nearest Neighbour Classifiers*, arXiv:2004.04523.
9. Leithardt, V. R. Q. (2020), "PRIPRO: A Comparison of Classification Algorithms for Managing Receiving Notifications in Smart Environments", *Applied Sciences*, 10(2), p. 502.
10. Andrić, I., Pina, A., Ferrão, P., Fournier, J., Lacarrière, B. and Le Corre, O. (2017), "Assessing the feasibility of using the heat demand-outdoor temperature function for a long-term district heat demand forecast", *Energy Procedia*, 116, pp. 460-469.
11. Pontil, M. and Verri, A. (1998), "Support vector machines for 3D object recognition", *IEEE transactions on pattern analysis and machine intelligence*, 20(6), pp. 637-646.
12. Muralidharan, R. and Chandrasekar, C. (2011), "Object recognition using support vector machine augmented by RST invariants", *International Journal of Computer Science Issues (IJCSI)*, 8(5), p. 280.
13. Bishop, C. M. (2006), *Pattern recognition and machine learning*, Springer.
14. Nabipour, M., Nayyeri, P., Jabani, H. and Mosavi, A. (2020), *Deep learning for Stock Market Prediction*, arXiv:2004.01497.
15. Skiena, S. S. (2017), *The data science design manual*, Springer.
16. Breiman, L. (2001), "Random forests", *Machine learning*, 45(1), pp. 5-32.
17. *RFC* (2020), available at: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
18. Chollet, F. (2018), *Deep Learning mit Python und Keras*, MITP-Verlags GmbH & Co. KG.
19. Yaloveha, V., Hlavcheva, D., Podorozhniak, A. and Kuchuk, H. (2019), "Fire hazard research of forest areas based on the use of convolutional and capsule neural networks", *2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, IEEE, pp. 828-832, DOI: <https://doi.org/10.1109/UKRCON.2019.8879867>
20. Kuchuk, H., Podorozhniak, A., Hlavcheva, D. and Yaloveha, V. (2020), "Application of Deep Learning in the Processing of the Aerospace System's Multispectral Images", *Handbook of Research on Artificial Intelligence Applications in the Aviation and Aerospace Industries*, IGI Global, pp. 134-147, DOI: [https://doi.org/10.4018/978-1-7998-1415-3\\_ch005](https://doi.org/10.4018/978-1-7998-1415-3_ch005)
21. Deng, F., Pu, S., Chen, X., Shi, Y., Yuan, T. and Pu, S. (2018), "Hyperspectral image classification with capsule network using limited training samples", *Sensors*, 18(9), pp. 3153.
22. LeCun, Y., Bengio, Y. and Hinton, G. (2015), "Deep learning", *Nature*, 521(7553), pp. 436-444.
23. *Google Colab* (2020), available at: [colab.research.google.com](https://colab.research.google.com)
24. Aksac, A., Demetrick, D. J., Ozyer, T. and Alhajj, R. (2019), "BreCaHAD: a dataset for breast cancer histopathological annotation and diagnosis", *BMC research notes*, 12(1), pp. 1-3.
25. Kim, J. (1997), *Iterated grid search algorithm on unimodal criteria*, Virginia Tech.
26. *Scikit learn* (2020), available at: <https://scikit-learn.org/stable/index.html>
27. Agarap, A. F. (2017), "An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification", arXiv preprint arXiv:1712.03541.
28. Hlavcheva, D., Yaloveha, V. and Podorozhniak, A. (2019), "Application of convolutional neural network for histopathological analysis", *Advanced Information Systems*, Vol. 3, No. 4, pp. 69-73, DOI: <https://doi.org/10.20998/2522-9052.2019.4.10>
29. de Souza, B. F., de Carvalho, A. C. and Soares, C. (2010), "A comprehensive comparison of ml algorithms for gene expression data classification", *The 2010 International Joint Conference on Neural Networks (IJCNN): IEEE*, pp. 1-8.

30. Dierks, T. (2007), "Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG)", *Journal of neuroscience methods*, 161(2), pp. 342-350.
31. *World Health Organization: Cancer* (2020), available at: [https://www.who.int/health-topics/cancer#tab=tab\\_1](https://www.who.int/health-topics/cancer#tab=tab_1)

Received (надійшла) 12.03.2020

Accepted for publication (прийнята до друку) 20.05.2020

#### ABOUT THE AUTHORS / ВІДОМОСТІ ПРО АВТОРІВ

- Главчева Дар'я Максимівна** – студентка кафедри обчислювальної техніки та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;  
**Daria Hlavcheva** – student of Computer Science and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;  
 e-mail: [agafina99@gmail.com](mailto:agafina99@gmail.com); ORCID ID: <http://orcid.org/0000-0001-6990-6845>
- Яловега Владислав Анатолійович** – асистент кафедри обчислювальної техніки та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;  
**Vladyslav Yaloveha** – Assistant Lecturer of Computer Science and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;  
 e-mail: [vladyslavyaloveha@gmail.com](mailto:vladyslavyaloveha@gmail.com); ORCID ID: <http://orcid.org/0000-0001-7109-9405>
- Подорожняк Андрій Олексійович** – кандидат технічних наук, доцент, доцент кафедри обчислювальної техніки та програмування, Національний технічний університет "Харківський політехнічний інститут", Харків, Україна;  
**Andrii Podorozhniak** – Candidate of Technical Sciences, Associate Professor, Associate Professor of Computer Science and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;  
 e-mail: [andriipodorozhniak@gmail.com](mailto:andriipodorozhniak@gmail.com); ORCID ID: <http://orcid.org/0000-0002-6688-8407>
- Лукова-Чуйко Наталія Вікторівна** – доктор технічних наук, доцент, доцент кафедри кібербезпеки та захисту інформації, Київський національний університет імені Тараса Шевченка, Київ, Україна;  
**Nataliia Lukova-Chuiko** – Doctor of Technical Sciences, Associate Professor, Associate Professor of Cyber Security and Information Protection Department, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine;  
 e-mail: [lukova@ukr.net](mailto:lukova@ukr.net); ORCID ID: <http://orcid.org/0000-0003-3224-4061>

#### Порівняння класифікаторів, у застосуванні до проблеми аналізу знімків біопсії

Д. М. Главчева, В. А. Яловега, А. О. Подорожняк, Н. В. Лукова-Чуйко

**Анотація. Мета дослідження** - порівняння алгоритмів класифікації для аналізу гістопатологічних зображень і оптимізація гіперпараметрів алгоритмів класифікації для збільшення результуючої точності. У статті вирішуються наступні **завдання**: попередньо обробленого набору даних BreCaHAD, реалізація та навчання CNN, застосування алгоритмів К-найближчих сусідів, SVM, Random Forest, XGBoost і перцептрона для класифікації ознак, отриманих CNN, і порівняння результатів. **Об'єктом дослідження** є процес класифікації пухлинних клітин на знімках мікроскопічної біопсії. **Предметом дослідження** є процес використання алгоритмів машинного навчання для класифікації ознак, отриманих CNN з вхідного зображення біопсії. **Наукова новизна** дослідження - порівняльний аналіз класифікаторів на завданні класифікації зображень «пухлинних» і «здорових» клітин з предоброботанного набору даних BreCaHAD. В результаті було отримано, що з обраних класифікаторів SVM має найбільшу точність на тестовій вибірці - 0,972. Це єдиний алгоритм, який отримав більшу точність, ніж перцептрон. Точність класифікації перцептрона на тестовій вибірці склала 0,966. К-найближчі сусіди, алгоритми Random Forest і XGBoost в результаті експериментів показали меншу точність. Проведена оптимізація гіперпараметрів алгоритмів. Результати були зіставлені з подібними роботами. Були використані такі методи дослідження: теорія глибокого навчання, математична статистика, оптимізація параметрів.

**Ключові слова:** машинне навчання; глибоке навчання; згортоква нейронна мережа; пошук по сітці; біопсія.

#### Сравнение классификаторов, в применении к проблеме анализа снимков биопсии

Д. М. Главчева, В. А. Яловега, А. А. Подорожняк, Н. В. Лукова-Чуйко

**Аннотация. Цель исследования** – сравнение алгоритмов классификации для анализа гистопатологических изображений и оптимизация гиперпараметров алгоритмов классификации для увеличения результирующей точности. В статье решаются следующие **задачи**: предварительная обработка изображений набора данных BreCaHAD, реализация и обучение CNN, применение алгоритмов К-ближайших соседей, SVM, Random Forest, XGBoost и перцептрона для классификации признаков, извлеченных CNN, и сравнение результатов. **Объектом исследования** является процесс классификации опухолевых клеток на снимках микроскопической биопсии. **Предметом исследования** является процесс использования алгоритмов машинного обучения для классификации признаков, извлеченных CNN из входного изображения биопсии. **Научная новизна** исследования - сравнительный анализ классификаторов на задаче классификации изображений «опухолевых» и «здоровых» клеток из предоброботанного набора данных BreCaHAD. В результате было получено, что из выбранных классификаторов SVM имеет наибольшую точность на тестовой выборке - 0,972. Это единственный алгоритм, который получил большую точность, чем перцептрон. Точность классификации перцептрона на тестовой выборке составила 0,966. К-ближайшие соседи, алгоритмы Random Forest и XGBoost в результате экспериментов показали меньшую точность. Проведена оптимизация гиперпараметров алгоритмов. Результаты были сопоставлены с подобными работами. Были использованы такие **методы исследования**: теория глубокого обучения, математическая статистика, оптимизация параметров.

**Ключевые слова:** машинное обучение; глубокое обучение; сверточная нейронная сеть; поиск по сетке; биопсия.