DOI 10.31891/2307-5732-2025-347-24
УДК 004.93

PAPCHENKO OLEKSANDR
Sumy State University, Sumy, Ukraine
https://orcid.org/0009-0005-9918-566X
e-mail: olpapchenko@gmail.com
BORYS KUZIKOV
Sumy State University, Sumy, Ukraine
https://orcid.org/0000-0002-9511-5665
e-mail: b.kuzikov@dl.sumdu.edu.ua

A HYBRID DEEP LEARNING AND INFORMATION-EXTREME APPROACH FOR BREAST CANCER HISTOPATHOLOGICAL IMAGE CLASSIFICATION

The paper presents a method for classifying breast cancer histopathological whole-slide images using a deep learning model based on ResNet-50 and an Information-Extreme Learning (IET) classifier, which is crucial considering the increasing mortality rate from breast cancer year after year. The gold standard method of breast cancer diagnosis based on histopathological whole-slide images (HI) requires manual inspection of every region of the high-resolution image, which is time-consuming and error-prone. Computer-aided systems (CAD) powered by machine learning can speed up the work of a histopathologist and reduce the error rate of decision-making, thereby improving the accuracy and efficiency of the diagnostic process. Analysis of existing solutions has shown that deep neural networks, especially convolutional neural networks (CNNs), have the highest rate of functional efficiency in image classification tasks due to their ability to learn and extract relevant features from the input data automatically. However, the training stage demands a large amount of high-quality data, which is often scarce in histopathology, making it an open task to develop high-performance models with limited training image datasets. The proposed model was trained on the publicly available BreakHis dataset, achieving an 88.98% accuracy after only 20 training epochs, outperforming other methods. The results demonstrate that using a pre-trained CNN on a non-domain-specific dataset can yield accurate feature extraction. This shows promise for enhancing CAD systems, especially in data-limited scenarios. The model was compared with simpler models based on local binary patterns for feature detection and support vector machines for classification, highlighting its superior performance and potential for real-world application in breast cancer diagnosis.

Keywords: convolutional neural networks, machine learning, transfer learning, breast cancer diagnosis, histopathology

ПАПЧЕНКО ОЛЕКСАНДР, КУЗІКОВ БОРИС
Сумський державний університет

ГІБРИДНИЙ ПІДХІД НА ОСНОВІ ГЛИБОКОГО НАВЧАННЯ ТА ІНФОРМАЦІЙНО-ЕКСТРЕМАЛЬНОЇ ТЕХНОЛОГІЇ ДЛЯ КЛАСИФІКАЦІЇ ГІСТОПАТОЛОГІЧНИХ ЗОБРАЖЕНЬ РАКУ МОЛОЧНОЇ ЗАЛОЗИ

У статті представлено метод класифікації гістопатологічних повнослайдових зображень раку молочної залози з використанням моделі глибокого навчання на основі ResNet-50 та класифікатора на основі інформаційно-екстремальної інтелектуальної технології (ІЕІТ), що є критично важливим з огляду на зростаючу щорічну смертність від раку молочної залози. Золотим стандартом діагностики раку молочної залози на основі гістопатологічних повнослайдових зображень є ручний огляд кожної області зображення високої роздільної здатності, що потребує значного часу та є схильним до помилок. Системи автоматизованої діагностики (CAD) на основі машинного навчання можуть прискорити роботу гістопатолога та зменшити частоту помилок при прийнятті рішень, тим самим підвищуючи точність та ефективність діагностичного процесу. Аналіз наявних рішень показав, що глибокі нейронні мережі, особливо згорткові нейронні мережі (CNN), мають найвищий показник функціональної ефективності в задачах класифікації зображень завдяки їхній здатності автоматично навчатися та виділяти релевантні ознаки з вхідних даних. Проте етап навчання потребує великого обсягу якісних даних, які часто обмежені в гістопатології, що робить актуальною задачу розробки високоефективних моделей з обмеженими наборами навчальних зображень. Запропонована модель була навчена на загальнодоступному наборі даних BreakHis, досягнувши точності 88,98% лише за 20 епох навчання, перевершуючи інші методи. Результати демонструють, що використання попередньо навченої CNN на наборі даних, не специфічному для конкретної області, може забезпечити точне виділення ознак. Це свідчить про перспективність підходу для вдосконалення CAD-систем, особливо в умовах обмеженості даних. Порівняння з іншими моделями продемонструвало її потенціал для практичного застосування в діагностиці раку молочної залози.

Ключові слова: згорткові нейронні мережі, машинне навчання, трансферне навчання, діагностика раку молочної залози, гістопатологія

Introduction

According to global cancer statistics, breast cancer is estimated to be the second leading cause of cancer death in women [1]. Early detection and treatment increase the chances of full recovery [2]. In Ukraine, the Chornobyl disaster led to a significant increase in breast cancer, especially in most contaminated areas [3]. With the war in Ukraine, the accessibility of medical services in general declined severely. In the case of breast cancer patients, it is essential to get diagnosed promptly; however, with the shortage of histopathologists caused by war actions on the territory of Ukraine, this is a challenging task [4]. In such a way, developing computer-aided systems for breast cancer detection and classification is in significant demand. Machine learning systems have already shown high functional efficiency in tasks of medical image analysis, in a study stating that AI/ML systems outperformed average radiologists by a significant margin (AUC improvement of +0.115, 95% confidence interval) [5], which paves application of such systems to other types of medical imaging.

1. Analysis of the subject area

1.1 Modern approaches to breast cancer diagnosis

Among the most widespread breast cancer diagnosis techniques are mammography, MRI, ultrasonography, PET, breast MI, and biopsy. Table 1 presents a comparison of different methodologies [6].

Table 1.

Comparison of breast cancer diagnosis methodologies

Method	Use	Limitations
Mammography	Mass screening. Scan soft tissue and blood vessels at the same time.	Ionizing radiation, effectiveness drops with tissue density.
Ultrasound	Evaluate lumps found during mammography.	Experienced operator required during examination.
MRI	High-risk young women can see small details of the tissue.	Some types of cancer are non-detectable using this method - such as ductal carcinoma.
CT	Detection of distant metastasis.	Radiation risk, expensive scanner.
PET	Functional imaging of biological process.	Ionizing radiation.
Biopsy	Histopathological whole-side imaging of the tissue is used to confirm diagnosis accurately.	Requires surgery intervention to collect tissue material.

Histological examination is considered the gold standard in cancer diagnosis [7]. Currently, it is almost the only way to confirm cancer diagnosis since noninvasive diagnosis methods cannot provide the same level of accuracy. In histopathology research, a specimen is dyed with stains (e.g., hematoxylin-eosin) to highlight tissue components under a microscope. According to the Nottingham Grading System, the assessment of breast cancer is mainly based on three morphological features in histology sections: tubule formation, nuclear pleomorphism, and the number of mitotic cells [8]. The whole side image is produced by photographing the tissue under a microscope. The output image has a high resolution, sometimes 100,000x100,000 pixels.

A huge resolution of the histological image computer-aided systems CAD is essential to make examination results faster and more reliable. On the one hand, the huge resolution makes it hard for operators to examine each image closely. Still, on the other hand, it contains huge amounts of high-quality data, making it possible to apply modern machine-learning approaches to image object detection.

1.2. Review of existing methods

CAD systems for histological image analysis have a long history of improvement and development [9]. Generally, ML methods applied in CAD systems can be divided into two main types: classical machine learning methods, which use manually defined image feature extraction techniques, and deep learning methods with automatic feature learning techniques [10]. Since 2012, a deep learning-based algorithm has outperformed classical image classification approaches, and there has been growing interest in developing deep learning-based systems [11]. The following section provides a brief overview of both approaches applied to the task of breast cancer detection by histological images.

Pin Wang et al, proposed automatic quantitative image analysis for breast cancer detection [12]. For the nuclei segmentation, top-bottom hat transform was applied, thus enhancing image quality. To obtain regions of interest (ROIs), wavelet decomposition and multi-scale region-growing (WDMR) were combined to obtain areas of interest (ROIs), thereby realizing precise location. Overlapped cells were split using a double-strategy model (DSSM) containing adaptive mathematical morphology and Curvature Scale Space (CSS) for better accuracy and robustness. To obtain optimal features, a support vector machine (SVM) with a chain-like agent genetic algorithm (CAGA). The proposed method used for testing 68 breast cancer histopathological images, classification reached high accuracy 96.19% (±0.31%) for accuracy, 99.05% (±0.27%) for sensitivity, and 93.33% (±0.81%) for specificity.

ML systems are generally data-driven, meaning the training dataset's quality directly influences the overall system functional efficiency in the exam mode. Spanhol et al. made available an online breast cancer histopathological images dataset containing 7909 images [13]. The original paper also contained approaches to solving the cancer image classification issue using hand-crafted feature descriptors. The authors used several texture descriptors: uniform Local Binary Pattern (LBP), Complete LBP (CLBP), Local Phase Quantization (LPQ), Grey Level Co-occurrence Matrix (GLCM), parameter-Free Threshold Adjacency Statistics (PFTAS), for classifiers were used four different classifiers: 1-Nearest Neighbor (1-NN), Quadratic Linear Analysis (QDA), Support Vector Machines (SVM), and Random Forests of Decision Trees. The accuracy was measured on the patient level and aggregated to get the overall score. PFTAS features vectors with SVM classifiers performed the best, having an overall accuracy of 85.1%. The dataset is widely used in research as a convenient tool for comparing the efficiency of different algorithms.

Using the same BreakHis dataset, Spanhol et al. developed a convolutional neural network (CNN) based system for cancer image classification [14]. Compared with the original paper for the BerakHis dataset [13], where hand-crafted feature descriptors were used, this paper used a deep learning approach for automatic feature learning. The paper also proposes the method for extraction and combination image patches used for the CNN training stage, thus decreasing the computational resource usage. Using patch sizes of 32×32 or 64×64 made it possible to use the AlexNet model [15]. Classification accuracy increased by around 6% compared with handcrafted feature descriptors.

Matos et al. continued the research based on the BreakHis dataset and image patching approach in their paper [15]. Their approach is based on patching the initial image - however, only relevantly classified patches are used for training. For discriminating patches as relevant and irrelevant, there was a separate SVM classifier trained on the CRC dataset [16] using as a feature descriptor in one case handcrafted PFTAS and in another pretrained InceptionV3 network [19]. Interestingly, pretraining of the InceptionV3 was made on the ImageNet dataset. On the second stage the classifier trained in the first stage was used to choose only relevant patches from BreakHis images to train the final SVM classifier. In such a way, there are two procedures of transfer learning, one from the CRC dataset to the BreakHis dataset and another one from ImageNet to the CRC and BreakHis datasets. The obtained system was able to classify images as malign and benign tumors with an accuracy of around 89% for the PFTAS features descriptor and around 97% accuracy for InceptionV3 as the features extractor.

In the work [18], an analysis of approaches used to diagnose breast cancer using machine learning techniques was conducted. The authors use a form of systematic review-structured evidence synthesis to compare different approaches. According to the review, algorithms based on deep learning and convolutional neural networks are state-of-the-art solutions that provide high functional efficiency for the end solution.

1.3. Conclusion on Existing Approaches

The dominance of ML methods based on the deep neural network is now evident. Universal neural network approximation theorem statements, together with modern techniques of image representation learning, make this class of methods one of the most powerful currently existing. New techniques of features embedding like attention [19], make the class of algorithm even more capable. However, this class of ML algorithms is highly data-driven, and the system accuracy depends on the quality of data provided during the training stage. Deep learning systems usually have a big learning capacity and are prone to overfitting. Developed numerous effective techniques to prevent model overfitting, for example, dropout [20]. Another problem is the need for a relatively big, balanced, diverse dataset. This is especially a problem in the field of histopathological cancer diagnosis. High-quality histological whole-side image datasets are hard to obtain. This is highly specialized imaging; usage of those assets needs legal approval from the patients. In such cases, especially important techniques like image augmentation [21] or synthetic dataset generation [22]. In conclusion, it's evident that deep learning-based algorithms have a proven level of accuracy. However, from the literature overview, it's clear that this approach cannot be easily transferred from one field to another; it requires the solution of numerous supplement tasks like neural network topology engineering, dataset preparation, features engineering, overfitting effective parameters learning, choosing the right cost function, etc.

2. Building a machine learning system

2.1. Dataset description

From the literature overview following the quality of training data directly affects the overall system accuracy in the exam mode. Thus, one of the first stages of ML system engineering is selection of the dataset. In our case, the BreakHis dataset of histopathological images [13] was chosen. The dataset has several advantages: it contains a large amount of labeled histopathological breast cancer images making it very widespread in the research field - which in turn makes it possible to compare our results with results of other approaches.

The dataset contains breast tumor microscopic images of several types. Labels contain tumor type as well as patient information. Samples are obtained from breast tissue biopsy slides, stained with hematoxylin and eosin (HE). The preparation routine is pretty standard, which makes this dataset transferable to real-life scenarios. Images were made in 3-channel RGB (Red-Green-Blue) TrueColor (24-bit color depth, 8 bits per color channel) color space having different magnifying factors of 40, 100, 200 and 400. Fig. 1 presented examples of dataset images with different magnification factors: (a) 40, (b) 100, (c) 200, and (d) 400.

The dataset contains a total of 7909 labeled images of a total of 82 patients. The dataset contains labels of benign and malignant types of cancer. However, it should be said that the dataset is not balanced - malignant classes have more items in the dataset. Table 2 presents image distribution by magnification and class.

Table 2

Presents image distribution by magnification and class

Magnification	Benign	Malignant	Total
40	625	1370	1995
100	644	1437	2081
200	623	1390	2013
400	588	1232	1820
Total	2480	5429	7909
Patients	24	58	82

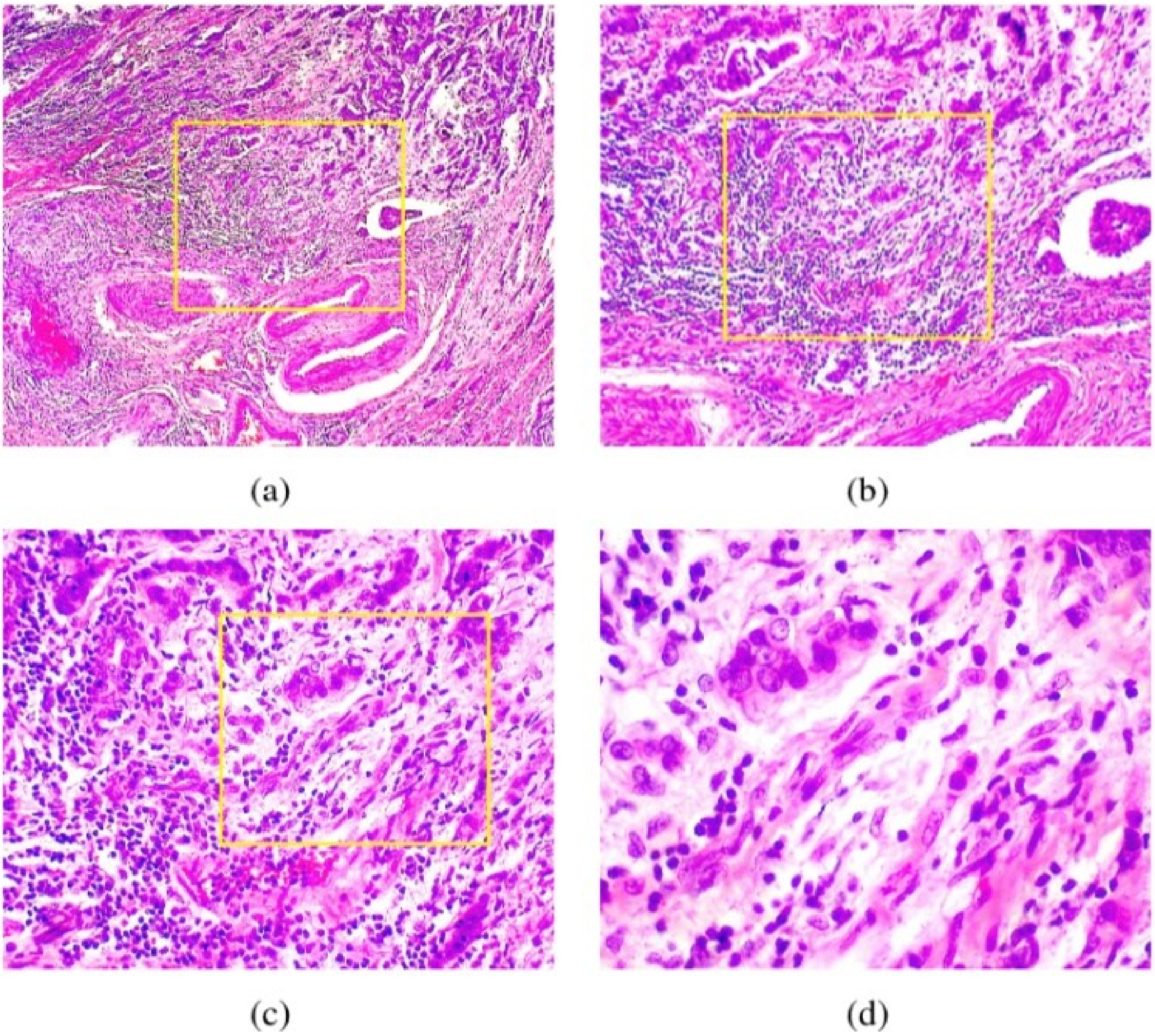

Fig. 1. BreakHis dataset images examples with different magnification factors (a) 40, (b) 100, (c) 200, and (d) 400. Highlighted rectangles are areas of interest selected by pathologists for magnification on the next image

2.2. Description of the proposed methods

The goal of the system is to perform binary classification of benign and malignant histopathological cancer images. For feature description, the proposed method is based on ResNet-50, which is pretrained on the ImageNet dataset [23]. ResNet-50 contains 50 layers. Network topology includes residual blocks - which make it possible to train deep neural networks effectively. Residual Blocks are skip-connection blocks that learn residual functions with reference to the layer inputs. Having such skip connections makes it easier to pass the data between layers of deep neural networks. As a result, gradient values are not dimmed at the top layers of the network, making the learning procedure faster. The classification layer of ResNet-50 was replaced with a global poling layer followed by a fully connected dense layer with a size of output 100. As activation of the dense layer was used, a rectified linear unit (RELU) [24]. For overfitting prevention, a dropout layer was used, with a dropout rate of 0.2. In such a way, a feature vector with a size of 100 can be obtained. For implementation, Tensorflow Keras API version 2.15.0 was used [25].

As a classifier of extracted features, the information extreme method (IET) was developed at Sumu State University [26]. The approach is based on the maximization of information criteria by finding optimal parameters in the process of training, thus forming an optimal information sense set of decisive rules. In our case, an IET approach with a second level of machine learning depth was used to optimize control tolerance and hyperspherical container radius.

The IEI approach performs transformation of the input matrix Y of training samples into training binary matrix X, which enables to perform with maximum possible probability correct classification decisions with the method of permissible transformations in machine learning. In the work [27], it was shown that the IET method's functional efficiency depends on the quality features of the engineering procedure. The Hamming binary space approach introduces a set of $\{g_m\}$ machine learning parameters which affect the functional efficiency of the IET algorithm. The set of optimal parameters has the following form:

$$g_m =< x_m, d_m, \delta_k >,$$

where x_m– is the average of the structured feature vector of the recognition class from the alphabet X_m^o; d_m – radius of the recognition class hyper spherical container for X_m^o, which is restored in the radial basis of the recognition features space; δ_k – parameter of the control tolerances for the recognition features field, which is equal to half of the symmetric field control tolerances.

As a criterion for optimization of machine learning parameters, IET is known to use the modified Kullback measure.

$$\boldsymbol{E_m^{(k)}} = \frac{\boldsymbol{n - \left(K_{1,m}^{(k)} + K_{2,m}^{(k)}\right)}}{\boldsymbol{n}} \frac{\boldsymbol{2n + \xi - K_{1,m}^{(k)} - K_{1,m}^{(k)}}}{\boldsymbol{K_{1,m}^{(k)} + K_{1,m}^{(k)} + \xi}}$$

where $K_{1,m}^{(k)}$ – number of false negative events, $K_{2,m}^{(k)}$ – number of false positive events, ξ – small number to prevent division on zero.

2.3. Training stage of the system

The ResNet-50-based backbone of the proposed approach was trained on the BreakHist dataset. Each image was preprocessed - converted from RGB format to BGR, then each color channel was zero-centered with respect to the ImageNet dataset, without scaling. To prevent ResNet-50 pretrained weights from being overwritten during the training stage, its weights were frozen, so that only dense layer weights were trained. The total number of dense layer trainable parameters is 204,900. As loss function was used binary cross entropy. A stochastic gradient based optimization approach was used - method based on adaptive estimates of lower-order moments [28]. Two separate runs of training were performed for 20 and 50 epochs. Actual training was performed on the Nvidia RTX 3050 GPU device. In Fig. 2, graphs of loss function per epoch are presented. As you can see from the graph the loss function value approximately after the 20th epoch is saturated and was not changing too much.

After the training on the input of the backbone, training samples were passed, as the input obtained feature vectors with a size of 100. The resulting feature vectors were used to train IET-based classifiers. For training, we used control tolerance ranging from 0 to 100. The training stage was performed for two sets of feature vectors - obtained as the result of the training backbone for 20 and 50 epochs, respectively. Fig. 3. presented graphs of information criterion relative value by control tolerance value for features from the two sets mentioned above. Information criterion's relative value is a fraction of max possible criterion value - reached on false positive and false negative values are zero. As can be seen from the graphs, the information criterion reached a bigger value for features obtained from backbone trained for 20 epochs at 0.78, compared to 50 epochs features at 0.46. Considering backbone training loss displayed in Fig. 2, we can assume that after the 20th epoch, our feature extraction dense layer started to overfit, thus producing worse results at the end.

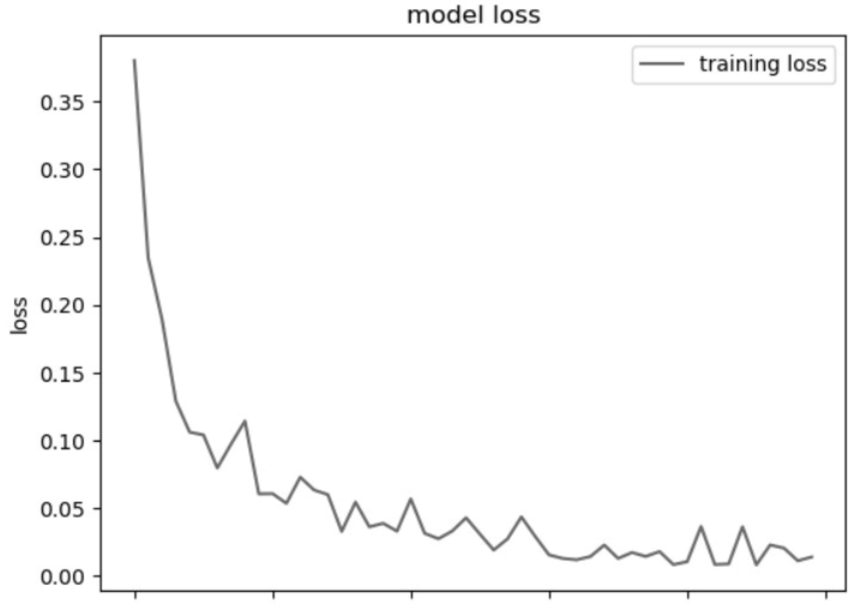

Fig. 2. Training binary cross entropy loss per epoch

The result of IET training was obtained following optimal parameters, see Table 3.

Table 3

IET optimal parameters

Backbone epochs	**Optimal control tolerance**	**Optimal Radii**	
		Malign	**Benign**
20	8	15	18
50	20	13	13

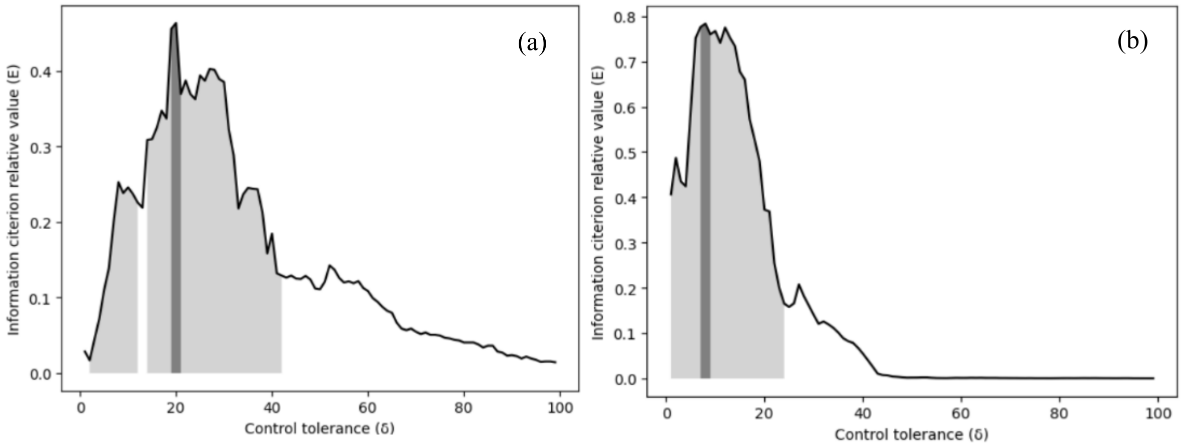

Fig. 3 Relative information criterion values by control tolerance values, light gray area - is working area, dark gray - is max value area, (a) - graph for IET training run using features from ResNet-50 based approach trained for 50 epochs, (b) - graph for IET training run using features from ResNet-50 based approach trained for 50 epochs.

3. Results

For accuracy evaluation of the obtained model was used the approach described in original BreakHis dataset paper [13]. The dataset was divided into training (70%) and testing subsets (30%). To guarantee the generalization of the classifier, the dataset was split into the patient-wise non-overlapping manner in training and test datasets - in such a way it is guaranteed that patients from the test set are not presented in the training set. For the actual split - was used script from the original paper [13] and [14] - making possible results comparison. For our model accuracy evaluation, images with a magnification of 200 only. We measured the accuracy image wise - meaning was measured total accuracy of the overall test set.

The test set contained a total of 744 samples. Results were measured separately for 20 and 50 epochs trained backbone features, see table 4.

Table 4

Accuracy results

Backbone epochs	**Accuracy**
20	88.98%
50	81.18%

As can be seen, a backbone that is trained for 20 epochs results in higher accuracy. We can also compare results for the approach from BreakHis dataset authors based on CNN. In their approach image, image-wise accuracy for 200 magnification factor was 84.0% with a standard deviation of 3.2 [14].

Conclusions

The problem of automatic classification of breast cancer as malignant and benign, based on computed histopathological images using machine learning algorithms, was considered.

Performed literature overview of existing approaches, compared their functional efficiency and existing challenges. The proposed approach to classifying breast cancer histopathological images based on multistage procedure- ResNet-50 based backbone for features extraction and IET classifier for binary classification. Based on the result of the research it can be stated that backbone pre-training even on non-domain specific dataset - like ImageNet produces good results for features extraction. This can be explained by the fact that spatial patterns and textures on the ImageNet and histopathological images of breast cancer still have common primitives. For example, we can imagine that features extracted from ImageNet grass image regions can have similar texture patterns with some tissue nuclei. This also explains a well-known fact why image augmentation increases the functional efficiency of image classification systems. We can assume that by using augmented images a network can learn more types of primitive patterns, textures and use it as building blocks for construction of more complicated units on the deep levels of the network. Similar results were obtained in research [14] where the AlexNet CNN network was pretrained on the ImageNet dataset. The obtained model performance has comparable results with state-of-the-art approaches.

The work was carried out as part of the Scientific Research Work "Information technology models and methods for analysis and synthesis of structural, information and functional models, objects and automated processes" (state registration number 0120U103071, Ministry of Education and Science of Ukraine)

References

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394–424.
2. Marmot, M. G., Altman, D. G., Cameron, D. A., Dewar, J. A., Thompson, S. G., & Wilcox, M. (2013). The benefits and harms of breast cancer screening: an independent review. *British Journal of Cancer*, 108(11), 2205–2240.

3. Pukkala, E., Kesminiene, A., Poliakov, S., Ryzhov, A., Drozdovitch, V., Kovgan, L., ... & Cardis, E. (2006). Breast cancer in Belarus and Ukraine after the Chernobyl accident. *International Journal of Cancer*, 119(3), 651–658. https://doi.org/10.1002/ijc.21885

4. Shushkevich, A. (2022). Inside the war: life in Ukraine. *International Journal of Gynecologic Cancer*, 32(5), 686–687. https://doi.org/10.1136/ijgc-2022-003556

5. McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Suleyman, M. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6

6. Wang, L. (2017). Early diagnosis of breast cancer. *Sensors*, 17(7), 1572. https://doi.org/10.3390/s17071572

7. Jothi, J. A. A., & Rajam, V. M. A. (2016). A survey on automated cancer diagnosis from histopathology images. *Artificial Intelligence Review*, 48(1), 31–81. https://doi.org/10.1007/s10462-016-9494-6

8. Lu, Y., Liu, A.-A., & Su, Y.-T. (2021). Mitosis detection in biomedical images. In *Computer Vision for Microscopy Image Analysis* (pp. 131–157). https://doi.org/10.1016/B978-0-12-814972-0.00006-0

9. Stenkvist, B., Bengtsson, E., Eriksson, O., Holmquist, J., Nordin, B., & Westman-Naeser, S. (1978). Computerized nuclear morphometry as an objective method for characterizing human cancer cell populations. *Cancer Research*, 38(12), 4688–4697.

10. Lee, H., & Chen, Y.-P. P. (2015). Image based computer aided diagnosis system for cancer detection. *Expert Systems with Applications*, 42(12), 5356–5365. https://doi.org/10.1016/j.eswa.2015.02.005

11. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (Vol. 25, pp. 1097–1105).

12. Wang, P., Hu, X., Li, Y., Liu, Q., & Zhu, X. (2016). Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Signal Processing*, 122, 1–13. https://doi.org/10.1016/j.sigpro.2015.11.011

13. Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7), 1455–1462. https://doi.org/10.1109/TBME.2015.2496264

14. Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). Breast cancer histopathological image classification using Convolutional Neural Networks. In *2016 International Joint Conference on Neural Networks (IJCNN)*. https://doi.org/10.1109/ijcnn.2016.7727519

15. De Matos, J., Pedrosa, J., & Lima, R. (2019). Double transfer learning for breast cancer histopathologic image classification. In *2019 International Joint Conference on Neural Networks (IJCNN)*. https://doi.org/10.1109/IJCNN.2019.8852092

16. Oliveira, S. P., Pereira, E. M., & Ramos, I. (2021). CAD systems for colorectal cancer from WSI are still not ready for clinical acceptance. *Scientific Reports*, 11(1). https://doi.org/10.1038/s41598-021-93746-z

17. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2016.308

18. Houssami, N., Kirkpatrick-Jones, G., Noguchi, N., & Lee, C. I. (2019). Artificial Intelligence (AI) for the early detection of breast cancer: a scoping review to assess AI's potential in breast screening practice. *Expert Review of Medical Devices*, 16(5), 351–362. https://doi.org/10.1080/17434440.201

19. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 6000–6010.

20. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.

21. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60. https://doi.org/10.1186/s40537-019-0197-0

22. Yang, W., & Nam, W. (2022). Data synthesis method preserving correlation of features. *Pattern Recognition*, 122, 108241. https://doi.org/10.1016/j.patcog.2021.108241

23. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

24. Xu, B., Wang, N., Chen, T., & Li, M. (2015). Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*.

25. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

26. Dovbysh, A. S., Klymenko, O. M., & Klymenko, M. O. (2023). Decision-making support system for diagnosis of breast oncopathologies by histological images. *Cybernetics and Systems Analysis*, 59(3), 493–502. https://doi.org/10.1007/s10559-023-00584-0

27. Papchenko, O., Kuzikov, B., & Shovkoplias, O. (2023). Enhancing functional efficiency in information-extreme machine learning with logistic regression ensembles. *Radioelectronic and Computer Systems*, (4), 65–74. https://doi.org/10.32620/reks.2023.4.06

28. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

29. Kingma D. P. Adam: a method for stochastic optimization [Electronic resource] / Diederik P. Kingma, Jimmy Ba. – Mode of access: https://doi.org/10.48550/ARXIV.1412.6980.