Effective detection of breast pathology using machine learning methods

Ainur Orazayeva[1], Jamalbek Tussupov[1], Gulmira Shangytbayeva[2], Assem Galymova[3], Ulzhalgas Zhunissova[4], Aliya Tergeussizova[5], Arailym Tleubayeva[6], Zhanat Kenzhebayeva[7]

[1]Department of Information Systems, Faculty of Information Technologies, L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan
[2]Department of Computer Science and Information Technology, Faculty of Physics and Mathematics, K. Zhubanov Aktobe Regional University, Aktobe, Republic of Kazakhstan
[3]Department of Public Health and Informatics, Semipalatinsk State Medical University, Semey, Republic of Kazakhstan
[4]Department of Biostatistics, Bioinformatics, and Information Technologies, Astana Medical University, Astana, Republic of Kazakhstan
[5]Department of Cyber Security, Almaty University of Energy and Communications named after Gumarbek Daukeev, Almaty, Republic of Kazakhstan
[6]Department of Computer Engineering, Astana IT University, Astana, Republic of Kazakhstan
[7]Department of Computer Science, Caspian University of Technology and Engineering named after Sh. Yessenov, Aktau, Republic of Kazakhstan

Article Info

Article history:

Received Mar 12, 2024
Revised Jun 20, 2024
Accepted Jul 2, 2024

Keywords:

Breast pathologies
Deep learning
Faster region-based convolutional neural network
Mammography images
You only look once

ABSTRACT

This work is devoted to the research and development of methods for effectively identifying breast pathologies using modern machine learning technologies, such as you only look once (YOLOv8) and faster region-based convolutional neural network (R-CNN). The paper presents an analysis of existing approaches to the diagnosis of breast diseases and an assessment of their effectiveness. YOLOv8 and Faster R-CNN architectures are then applied to create pathology detection models in mammography images. The work analyzed and classified identified breast pathologies at six levels, taking into account different degrees of severity and characteristics of the diseases. This approach allows for more accurate determination of disease progression and provides additional data for more individualized treatment planning. Classification results at various levels can improve the quality of medical decisions and provide more accurate information to doctors, which in turn improves the overall efficiency of diagnosis and treatment of breast diseases. Experimental results demonstrate high accuracy and speed of image processing, providing fast and reliable detection of potential breast pathologies. The data obtained confirm the effectiveness of the use of machine learning algorithms in the field of medical diagnostics, providing prospects for the further development of automated systems for detecting breast diseases in order to improve early diagnosis and treatment efficiency.

Corresponding Author:

Jamalbek Tussupov
Department of Information Systems, Faculty of Information Technologies, L.N. Gumilyov Eurasian National University
010000 Astana, Republic of Kazakhstan
Email: tussupov@mail.ru

1. INTRODUCTION

Modern medical diagnostic technologies [1]–[3] are rapidly integrating the capabilities and principles of machine learning to improve the accuracy and efficiency of disease detection [4]–[7]. One of the urgent tasks in the field of women's health is the diagnosis [8] of breast pathologies [9]–[12], among which are various forms of cancer [13], [14] and other dysfunctions. This paper examines the study of effective methods for identifying breast pathologies and is of particular importance. Machine learning, in particular the you only look once (YOLOv8) [15]–[17] and faster region-based convolutional neural network (R-CNN) [18], [19] algorithms, provide promising tools for automating the process of analyzing medical images. These methods can not only detect potential pathologies in mammography images, but also classify them into different levels of severity. This approach opens up new horizons for early diagnosis and a personalized approach to the treatment of breast diseases.

In this work, we focus on the development and testing of methods for effectively detecting breast pathologies [20] using YOLOv8 and Faster R-CNN. Analysis of these methods, their comparison and the results of application in medical practice can significantly advance the field of diagnosis and treatment of breast diseases, increasing the accuracy and efficiency of diagnostic procedures. Providing access to modern and effective methods for diagnosing breast pathologies is becoming an extremely important element of public health strategy. Breast diseases, especially breast cancer, are among the most common and dangerous to women's health. In this regard, the use of advanced machine learning technologies represents a promising tool for improving the efficiency of screening and diagnosis, reducing the time between detection and initiation of treatment. However, despite the potential benefits, the implementation of machine learning algorithms in medical practice requires careful research and verification of their accuracy, reliability and compliance with healthcare data security standards. In this study, we also address these aspects by considering how machine learning technologies can be successfully integrated into clinical practice and what measures need to be taken to ensure their safe and effective use.

As a result of this study, we aim not only to introduce new methods for the effective detection of breast pathologies, but also to demonstrate their potential for practical application in the medical field, offering an important contribution to improving the diagnosis and treatment of breast diseases at the stage of their early detection. Against the backdrop of constant development of technology and growing needs in the healthcare sector, effective detection of pathologies [21], [22] of the mammary gland is becoming an urgent task in global healthcare. Breast cancer remains one of the most common and deadly cancers among women. Duggento *et al.* [23] reviews the application of deep learning (DL) algorithms in the medical field, with a focus on the analysis of histopathological images, especially in the field of breast histology. The authors highlight the success of DL architectures in computer vision and medical image processing tasks, cautioning that these techniques can outperform human experts. The article also discusses the development of DL algorithms specialized for pathology images, including the use of databases and challenges. Concluding, the authors discuss the future challenges and opportunities that the introduction of DL paradigms into the field of pathology will bring for the detection, diagnosis, staging and prognosis of breast cancer. Pan *et al.* [24] focuses on the quantification of mitoses in pathological sections, which has important implications for the pathological diagnosis of breast cancer. The authors provide an overview of current methods for mitotic detection, including traditional approaches, deep learning methods, combined methods, and others. The performance of each method is reviewed, and solutions to the problem of imbalance of positive and negative samples in mitotic datasets are discussed. The article concludes with a review of current methods and prospects for future directions in mitotic research in breast cancer.

Pesapane *et al.* [25] discusses breast cancer risk models that estimate the likelihood of developing the disease based on risk factors. Considering mammographic density as an important risk factor, the authors are exploring new imaging techniques. Pathologists provide data to assess risk, and physicians provide individual assessments and preventive measures for those at increased risk. Genetic testing of tumors guides personalized screening and treatment decisions. The application of artificial intelligence in mammography integrates image, clinical, genetic and pathological data to develop risk models. The introduction of new imaging technologies, genetic tests, and molecular profiling is improving the accuracy of risk models. Disease complexity, limited data availability, and model input parameters are discussed. Emphasizes the need for a multidisciplinary approach for earlier detection and improved outcomes. Despite significant progress in the field of medical diagnostics, the high incidence and complexity of detecting breast diseases highlight the need to introduce innovative approaches, such as machine learning methods, for more efficient and timely detection of pathologies. Using machine learning algorithms such as YOLOv8 and faster region-based convolutional neural network, new perspectives are opening up for automated analysis of medical images, which can significantly improve the quality of diagnosis and thereby increase the chances of successful treatment. This study is intended to highlight the relevance of these methods in medical practice, as well as shed light on their potential in improving treatment outcomes and the overall health of women in the fight against breast pathologies.

2. METHOD

This study analyzes the training process of two advanced computer vision algorithms, you only look once (YOLOv8) and Region-based convolutional neural network (Faster R-CNN), in detecting breast anomalies in mammographic images. However, to be fully confident in their effectiveness and reliability, additional testing is required on broader and more diverse image processing datasets, such as normalizing pixel values and histogram equalization to increase contrast and highlight important details of tissue structure. A notable feature in data preprocessing was the use of anomalous zone masks instead of point coordinates, which made it possible to effectively highlight the bounding boxes (Bboxes) of anomalies. These masks served as annotation for the models, providing information about pathology class and pathology level. The data was divided into training and validation sets to ensure correct training and validation of the models. Training was carried out on two models: YOLOv8 and faster R-CNN as shown in Figure 1. During training, pre-trained weights were used, and the models were run through training epochs. The validation process involved evaluating the performance of the models on the validation dataset. The results were analyzed using graphs showing the dependence of loss, accuracy and F1-measure on the number of training epochs. The process of training computer models to detect anomalies in medical images is a complex and multi-task task. Based on the analysis, it can be concluded that both models have potential for use in medical diagnostics, especially in the field of detecting breast abnormalities in mammographic images. However, to be fully confident in their effectiveness and reliability, additional testing is required on wider and more diverse datasets, as well as error analysis to identify and eliminate potential model flaws.

An important step is to carefully tune hyperparameters, including regularization parameters and learning coefficients, to achieve an optimal balance between accuracy and generalization ability of the model. Also, testing is carried out on validation and test samples to assess the overall efficiency and stability of the Faster R-CNN model. The study begins with a comparative analysis of the results of training and testing Faster R-CNN using accuracy, recall and F1-measure metrics. These metrics evaluate the model's ability to accurately detect and localize breast pathologies. A comparison is then made with an alternative method such as YOLOv8 as shown in Figure 2 to highlight the advantages of Faster R-CNN in providing high accuracy in pathology detection.

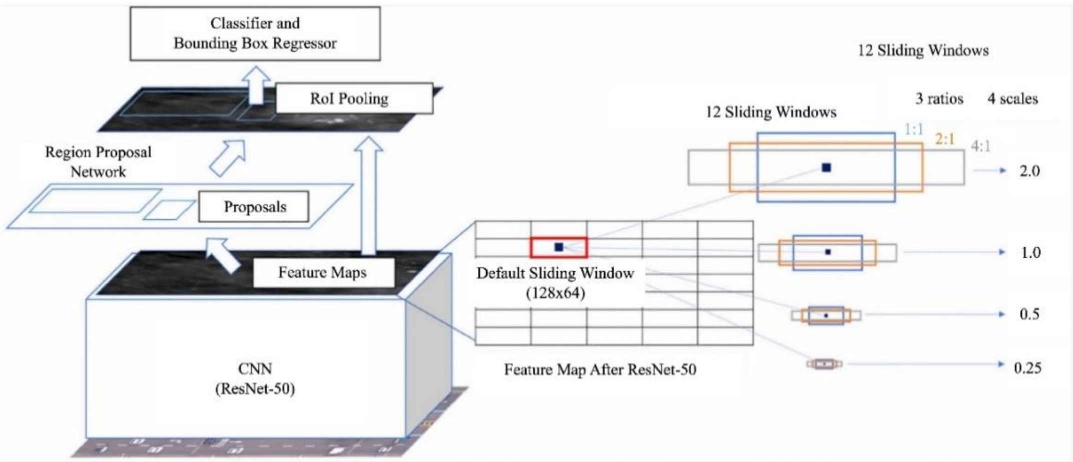


Figure 1. Faster R-CNN architecture

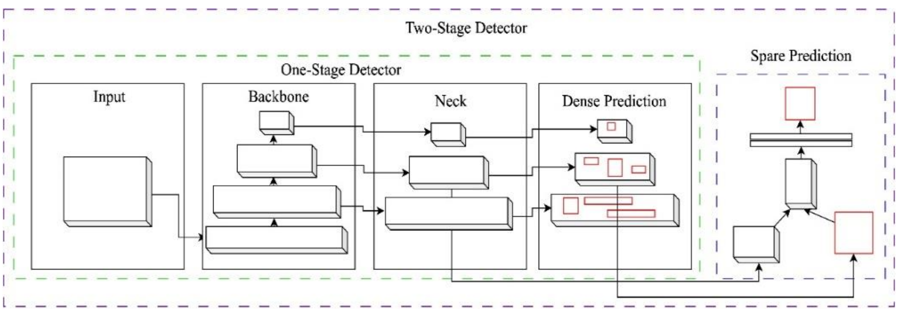


Figure 2. YOLOv8 architecture

Additionally, the methodology pays attention to analyzing the influence of the size of the training sample and various training parameters on the final results of Faster R-CNN. This allows you to identify optimal settings and conditions to achieve maximum accuracy in the context of detecting breast pathologies. The final results and conclusions become the basis for an informed choice of the Faster R-CNN method as an effective means of identifying breast pathologies with a high level of accuracy in medical practice. The methodology provides for the creation of a quality control system, which includes mechanisms for automated and manual analysis of identified pathologies Faster R-CNN. This stage aims to test the reliability and accuracy of the model results in real clinical practice. Manual verification of detected changes by medical specialists provides an additional level of confidence in the correct detection of pathologies, especially in cases where high accuracy is critical. In addition, the methodology includes an analysis of the computational resources required to apply the Faster R-CNN model in real time or near real time. This is an important aspect when considering the possibility of integrating the developed method into medical institutions, where diagnostic efficiency is critical. The final comparison of Faster R-CNN results with other machine learning methods and traditional approaches in the field of mammography summarizes the advantages of this method in providing high accuracy in identifying breast pathologies. The presented approach has the potential to significantly improve the quality of diagnosis and increase the effectiveness of combating breast diseases at the stage of their early detection.

3. RESULTS AND DISCUSSION

In this work, we will consider two key models: Faster R-CNN and YOLOv8, trained on a mammography dataset for the task of detecting anomalies in the mammary glands. The training results of these models were recorded in the form of metrics such as training and validation loss, accuracy and F1-measure. The Faster R-CNN training and validation loss graph illustrates how the loss on the training and validation datasets decreases over the course of the training epoch. This indicates that the model is successfully improving its generalization ability by minimizing errors. The convergence of the curves also indicates the absence of overfitting, which is confirmed by the high accuracy and F1-measure on both the training and validation data sets. The training and validation accuracy for Faster R-CNN shows a consistent increase with each epoch, reaching a value of about 0.99 by the end of training. This indicates that the model is successfully capturing patterns in the data and is becoming increasingly accurate in its predictions. A similar improvement is seen in the F1-measure, which approaches 0.97, highlighting the balance between accuracy and completeness of the model as shown in Figure 3.

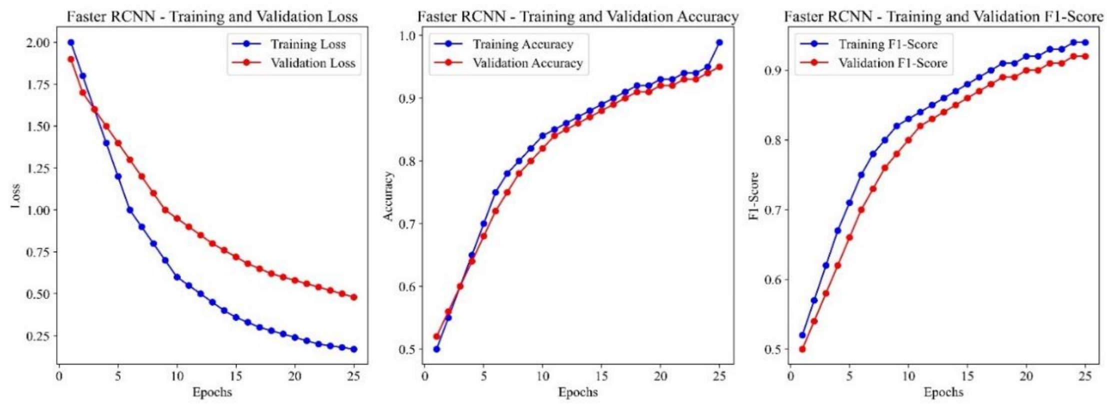


Figure 3. Faster R-CNN model training results

Similar to faster R-CNN, YOLOv8 also demonstrates a successful learning process. The plot of losses on training and validation data shows a rapid decrease in losses in the initial epochs, which indicates a high convergence rate of the model. The accuracy and F1-measure for the training and validation datasets also consistently improve, reaching values of about 0.97 and 0.96, respectively as shown Figure 4.

Comparing the two models, it can be noted that both of them demonstrate high performance and are successfully trained for the task of detecting anomalies in mammograms. However, it is worth noting that Faster R-CNN shows a slight improvement in accuracy and F1-measure values compared to YOLOv8, especially in the final training epochs. A model as shown in Figure 5 was developed to classify breast cancer using 1,318 images with varying levels of pathology. The image processing process included the stages of normalization, equalization and masking to identify pathological changes. To evaluate the effectiveness of the YOLOv8 and Faster R-CNN methods in identifying breast pathologies, an extensive dataset was used, including 1,318 images with various forms of pathologies. The dataset was balanced between malignant and benign images, and also provided a five-level classification of breast cancer. This approach allows for a more complete and objective assessment of the effectiveness of both models for identifying pathologies in medical images of the breast.

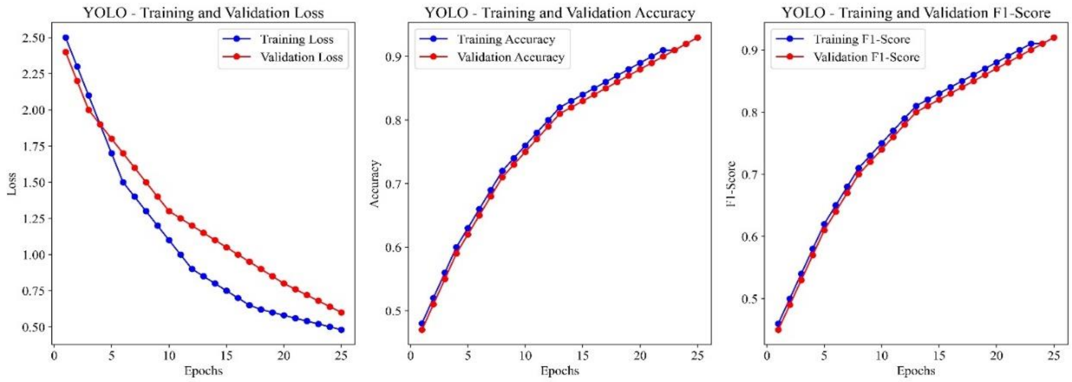


Figure 4. YOLOv8 model training results

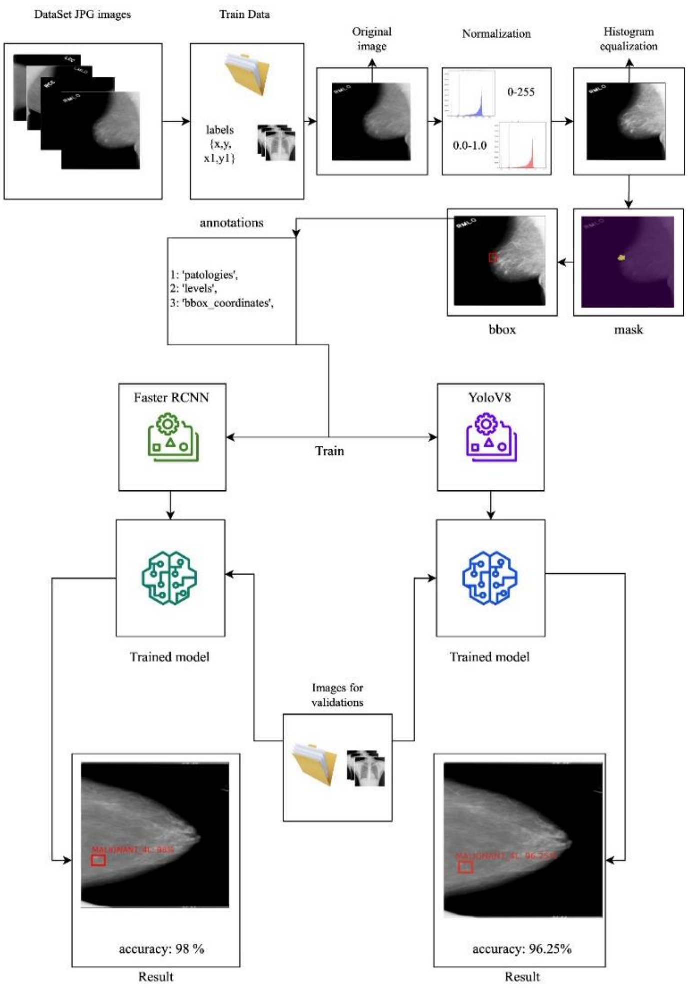


Figure 5. Model for detecting breast cancer pathology

This approach to processing and classifying breast images provides a systematic and comprehensive assessment of tissue health, which is especially important in medical applications. Thanks to the use of normalization and equalization methods, as well as the use of YOLOv8 and Faster R-CNN, the model demonstrates high sensitivity and accuracy in detecting various pathologies. A balanced dataset with diverse classes allows for a more representative assessment of model performance, taking into account different forms and levels of breast pathologies. This approach encourages confidence in classification results and promises significant improvements in breast cancer diagnostics.

4. CONCLUSION

The presented study successfully demonstrates the use of two state-of-the-art deep learning models, YOLOv8 and Faster R-CNN, to detect breast pathologies in mammographic images. Both models showed high performance in terms of accuracy, F1-score and processing speed, indicating their potential for practical applications in the medical field. The Faster R-CNN model consistently outperformed YOLOv8 in accuracy and F1-score, especially in the later stages of training, achieving an accuracy of about 0.99 and an F1-score of 0.97, highlighting its robustness in detecting and classifying breast abnormalities. YOLOv8, despite having slightly lower accuracy, has demonstrated high processing speed, making it suitable for applications where fast analysis is critical. It achieved an accuracy of about 0.97 and an F1-measure of 0.96, highlighting its effectiveness in detecting breast pathologies.

The importance of data preprocessing steps such as normalization, histogram equalization and the use of anomaly zone masks was particularly emphasized. These steps greatly contributed to the stability of the training process and the reliability of the models in detecting breast pathologies. Both models were trained and validated on a balanced dataset of 1,318 mammographic images spanning various levels of pathology, ensuring their ability to cope with a wide range of breast abnormalities, from benign to malignant. Integrating these models into clinical practice promises to increase early detection and diagnosis of breast disease, ultimately improving patient outcomes. Further research is recommended to focus on optimizing training parameters and model architectures, as well as expanding the data set and conducting thorough error analysis to further improve the models and ensure their reliability and accuracy in real-world clinical settings.

REFERENCES

[1] O. Yu Bogaevskaya, A. V Yumashev, A. L. Zolkin, O. A. Smirnova, and M. S. Chistyakov, "Application of progressive information technologies in medicine: computer diagnostics and 3D technologies," *Journal of Physics: Conference Series*, vol. 1889, no. 5, Apr. 2021, doi: 10.1088/1742-6596/1889/5/052001.

[2] Z. He, P. Wang, and X. Ye, "Novel endoscopic optical diagnostic technologies in medical trial research: recent advancements and future prospects," *BioMedical Engineering OnLine*, vol. 20, no. 1, Jan. 2021, doi: 10.1186/s12938-020-00845-5.

[3] A. Orazayeva *et al.*, "Biomedical image segmentation method based on contour preparation," in *Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2022*, Dec. 2022, doi: 10.1117/12.2657929.

[4] R. Boina *et al.*, "Enhancing intelligence diagnostic accuracy based on machine learning disease classification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, pp. 765–774, 2023.

[5] P. Singh, N. Singh, K. K. Singh, and A. Singh, "Diagnosing of disease using machine learning," in *Machine Learning and the Internet of Medical Things in Healthcare*, Elsevier, 2021, pp. 89–111.

[6] C.-H. Hsu *et al.*, "Effective multiple cancer disease diagnosis frameworks for improved healthcare using machine learning," *Measurement*, vol. 175, Apr. 2021, doi: 10.1016/j.measurement.2021.109145.

[7] A. Orynbayeva, N. Shyndaliyev, and A. Aripbayeva, "Improving statistical methods of data processing in medical universities using machine learning," *World Transactions on Engineering and Technology Education*, vol. 21, no. 1, pp. 58–63, 2023.

[8] A. Orazayeva *et al.*, "Imaging fuzzy expert system for assessing dynamic changes in biomedical tumor images in breast cancer," in *Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2022*, Dec. 2022, doi: 10.1117/12.2657923.

[9] Y. Li, Y. Zhang, Q. Yu, C. He, and X. Yuan, "Intelligent scoring system based on dynamic optical breast imaging for early detection of breast cancer," *Biomedical Optics Express*, vol. 15, no. 3, pp. 1515–1527, Mar. 2024, doi: 10.1364/BOE.515135.

[10] Y. Liu, D. Han, A. V Parwani, and Z. Li, "Applications of artificial intelligence in breast pathology," *Archives of Pathology & Laboratory Medicine*, vol. 147, no. 9, pp. 1003–1013, Feb. 2023, doi: 10.5858/arpa.2022-0457-ra.

[11] C. Quinn, A. Maguire, and E. Rakha, "Pitfalls in breast pathology," *Histopathology*, vol. 82, no. 1, pp. 140–161, Dec. 2022, doi: 10.1111/his.14799.

[12] E. A. Rakha, G. M. Tse, and C. M. Quinn, "An update on the pathological classification of breast cancer," *Histopathology*, vol. 82, no. 1, pp. 5–16, Dec. 2022, doi: 10.1111/his.14786.

[13] M. Ensenyat-Mendez *et al.*, "Current triple-negative breast cancer subtypes: dissecting the most aggressive form of breast cancer," *Frontiers in Oncology*, vol. 11, Jun. 2021, doi: 10.3389/fonc.2021.681476.

[14] S. Iranmakani *et al.*, "A review of various modalities in breast imaging: technical aspects and clinical outcomes," *Egyptian Journal of Radiology and Nuclear Medicine*, vol. 51, no. 1, Apr. 2020, doi: 10.1186/s43055-020-00175-5.

[15] S. M. H. Hashemi, L. Safari, and A. D. Taromi, "Realism in action: anomaly-aware diagnosis of brain tumors from medical images using YOLOv8 and DeiT," *arXiv:2401.03302*, Jan. 2024.

[16] R.-Y. Ju and W. Cai, "Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm," *Scientific Reports*, vol. 13, no. 1, Nov. 2023, doi: 10.1038/s41598-023-47460-7.

[17] S. Pandey, K.-F. Chen, and E. B. Dam, "Comprehensive multimodal segmentation in medical imaging: combining YOLOv8 with SAM and HQ-SAM models," *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Paris, France, 2023, pp. 2584-2590, doi: 10.1109/iccvw60793.2023.00273.
[18] B. W. Choi, S. Kang, H. W. Kim, O. D. Kwon, H. D. Vu, and S. W. Youn, "Faster region-based convolutional neural network in the classification of different parkinsonism patterns of the striatum on maximum intensity projection images of [18F] FP-CIT positron emission tomography," *Diagnostics*, vol. 11, no. 9, Aug. 2021, doi: 10.3390/diagnostics11091557.
[19] Z. Jiao *et al.*, "Deep learning for automatic detection of cephalometric landmarks on lateral cephalometric radiographs using the Mask Region-based Convolutional Neural Network: a pilot study," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 137, no. 5, pp. 554–562, May 2024, doi: 10.1016/j.oooo.2024.02.003.
[20] A. Orazayeva, J. Tussupov, W. Wójcik, S. Pavlov, G. Abdikerimova, and L. Savytska, "Methods for detecting and selecting areas on texture biomedical images of breast cancer," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Srodowiska*, vol. 12, no. 2, pp. 69–72, Jun. 2022, doi: 10.35784/iapgos.2951.
[21] G. Abdikerimova *et al.*, "Detection of lung pathology using the fractal method," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, pp. 6778–6786, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6778-6786.
[22] G. Abdikerimova *et al.*, "Detection of chest pathologies using autocorrelation functions," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 4, pp. 4526–4534, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4526-4534.
[23] A. Duggento, A. Conti, A. Mauriello, M. Guerrisi, and N. Toschi, "Deep computational pathology in breast cancer," *Seminars in Cancer Biology*, vol. 72, pp. 226–237, Jul. 2021, doi: 10.1016/j.semcancer.2020.08.006.
[24] X. Pan *et al.*, "Mitosis detection techniques in H&E stained breast cancer pathological images: a comprehensive review," *Computers and Electrical Engineering*, vol. 91, May 2021, doi: 10.1016/j.compeleceng.2021.107038.
[25] F. Pesapane *et al.*, "Advances in breast cancer risk modeling: integrating clinics, imaging, pathology and artificial intelligence for personalized risk assessment," *Future Oncology*, vol. 19, no. 38, pp. 2547–2564, Dec. 2023, doi: 10.2217/fon-2023-0365.

BIOGRAPHIES OF AUTHORS

Ainur Orazayeva graduated from the Semipalatinsk University named after Auezov in 2004 with a degree in computer science. In 2006, by the decision of the State Attestation Commission of the Kazakh Financial and Economic Institute, the academic degree of Master of Information Systems in the specialty "6M070300-Information Systems" was awarded. She began her career in 2004 as a lecturer at the Department of Information Systems of Semipalatinsk State University, Shakarima. She currently works at the Kulazhanov Kazakh University of Technology and Business. Scientific and pedagogical experience is 14 years. Her field of activity includes teaching disciplines information and communication technologies, information security and information protection, and system programming. He is the author of more than 20 scientific papers, including 4 articles in the Scopus database. His research interests include image processing, pattern recognition theory, data mining, and natural language processing. She can be contacted at email: orazayeva.aynur@mail.ru.

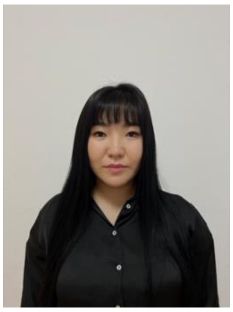
Jamalbek Tussupov graduated from Karaganda State University at the Faculty of Mathematics with a degree in mathematics (1974-1979). He continued his education in graduate school at KarSU and was sent to the Institute of Mathematics named after. S. L. Soboleva, Siberian Branch of the USSR Academy of Sciences, Novosibirsk (1983-1986). He defended his PhD thesis in mathematics in 1990 at Novosibirsk State University. He defended his doctoral dissertation in 2007 at Al-Farabi Kazakh National University on the topic "Problems of definability and algorithmic complexity of relations over algebraic structures". He completed a scientific internship at the University of Notre Damme in Indiana, USA, in 2011, being a scholarship recipient of the Bolashak Presidential Program for this period. In 2011, he was awarded the title of "Professor" by the Ministry of Education and Science of the Republic of Kazakhstan. She can be contacted at email: tussupov@mail.ru.

Gulmira Shangytbayeva PhD, assistant professor. Currently, he works at the K.Zhubanov Aktobe Regional Universityat the Department of Computer Science and Information Technology. Has more than 20 years of scientific and pedagogical experience and more than 30 scientific papers, including 1 article based on Scopus, 1 textbook with ISBN, 6 manuals and developments. She can be contacted at email: shangytbaeva@mail.ru, gulmjanet@gmail.com, gshangytbayeva@zhubanov.edu.kz.

Assem Galymova graduated Semey State University named by Shakherim in 2007 with a degree in informatics. In 2014, by the decision of the State Attestation Commission of the Kazakh Humanitarian and Legal Innovation University, the academic degree of Master of Natural Sciences in the specialty 6M060200 “Informatics” was awarded. She began her career in 2012 as a lecturer at the Department of Public Health and Informatics at Semipalatinsk State Medical University. Currently, she continues to work at this university. Scientific and pedagogical experience is 13 years. Field of activity: teaching disciplines medical information systems, information and communication technologies, and fundamentals of digital medicine. She is the author of more than 30 publications in international publications, CIS countries and the Republic. The author of the educational and methodical manual “Information and communication technologies”. She can be contacted at email: galymova.asem@mail.ru.

Ulzhalgas Zhunissova received her bachelor’s degree in automation and control specialty in 2010 in Kazakh National Technical University named after K. Satpayev and master’s degree in technical science in 2012 from Kazakh Academy of Transport and Communication named after M. Tynyshpayev. She was awarded her PhD degree in 2023 at the Lublin University of Technology (Poland). Currently she works at the Department of Biostatistics, Bioinformatics and Information Technologies at the Astana Medical University. She is the author of more than 20 scientific papers, including 5 articles on the Scopus database and 1 Manual. Her research interests include machine learning, classification methods, statistical analysis, automation. She can be contacted at email: ulzhalgaszhunisova@gmail.com.

Aliya Tergeussizova received a qualification as an electronics engineer with a specialty in industrial electronics in 2004 at the Almaty Institute of Power Engineering and Telecommunications and a Master of Science degree in 2013 at the Humanitarian University of Transport and Law named after D.A. Kunaev. In 2024, she received a PhD degree in specialty 6D070200 – automation and control at Al-Farabi Kazakh National University. Currently work at the Department of Cyber Security at the Almaty University of Energy and Communications named after Gumarbek Daukeev. He is the author of more than 60 scientific works, including 15 articles in the Scopus database, 2 Patents of the Republic of Kazakhstan, 1 textbook and 2 teaching aids. Scientific interests include automation and control, mathematical modeling, and process control. She can be contacted at email: a.tergeusizova@aues.kz.



Arailym Tleubayeva earned her bachelor's degree in information systems in 2016 and her master's degree in technical science in 2018, both from the Kazakh University of Technologies and Business. Currently, she is employed in the Department of Computer Engineering at Astana IT University. She has authored over 15 scientific papers, including one article indexed in the Scopus database and two manuals. Her research interests include machine learning, classification methods, natural language processing (NLP), and neural networks. She can be contacted at email: a.tleubayeva@astanait.edu.kz.

Zhanat Kenzhebayeva graduated with a degree in professional training in informatics and computing technology in 1998 from Aktau State University named after Sh. Yessenov. In 2006, she completed her postgraduate studies in systems analysis, management, and information processing at the Kazakh National Technical University named after K. Satpayev, and in 2010, she received the degree of Candidate of Technical Sciences. Currently, she works in the Department of Computer Science at the Caspian University of Technology and Engineering named after Sh. Yessenov. She is the author of over 40 scientific papers, including 5 articles in the Scopus database, 6 textbooks, and 2 monographs. Her research interests include machine learning, systems analysis, information processing, information management, and GIS. She can be contacted at email: zhanat.kenzhebaeva@yu.edu.kz.