



**Vilnius
University**

Enhancing Breast Cancer Recognition using Vision Transformer and Explainable AI

Supervisor: Prof. Povilas Treigys
Doctoral Student: Sobia Dastgeer
Academic year: 2025/2026
Doctoral study period: 2024-2028

CONTENT

1. Study plan and summary
2. Conferences & Publications
3. International Events and Publications
4. Exams
5. Research object and aim
6. Tasks
7. Stages of research and dissertation preparation
8. Scientific results obtained during the six-month period
9. Workplan for next half year
10. Study results

Vilnius University

Year of study	Participation in Conferences				Publications					
	International		National		With citation indicator			Without citation indicator		
	Plan	Completed	Plan	Completed	Plan	Completed	Condition	Plan	Completed	Condition
I (2024/2025)			2	2				0	1	Published
II (2025/2026)	1	1						2		
III (2026/2027)	1				1					
IV (2027/2028)	1				1					
Total:	3	1	2	2	2			2	1	

The plan for the reporting half-year and its implementation:

Exams 2025/2026 spring semester(4 th semester)		
Plan	Completion Date	Condition
Deep Neural Networks	2026 Q3	Pass

General skills 2025/2026 spring semester(4 th semester)	
Plan	Date
The 22nd Reasoning Web Summer School, Vilnius Lithuania	27–30 August 2026

Internship at foreign Institute		
Plan	Start Date	End Date
The AGH University of Krakow, Faculty of Computer Science, Poland	16-02-2026	15-05-2026

The plan for the reporting half-year and its implementation:

Conference participation 2025/2026 spring semester(4 th semester)		
Plan	Completed	Conference type
1	Authors: Dastgeer, S., Treigys, P. Poster Title: An Interpretable Attention-Guided Hybrid Deep Learning Framework for Breast Cancer Classification Conference name: 7th Polish Conference on Artificial Intelligence Date: April 20-22, 2026 Location: Krakow, Poland	International

Publications 2025/2026 spring semester(4 th semester)			
Plan	Completed	Condition	Publication Type

Information about international events and publications presenting the main results of the dissertation (articles with citation index only)

Participation in international conferences	
	Description Authors: Dastgeer, S., Treigys, P. Poster Title: An Interpretable Attention-Guided Hybrid Deep Learning Framework for Breast Cancer Classification Conference name: 7th Polish Conference on Artificial Intelligence Date: April 20-22, 2026 Location: Krakow, Poland
Publications (with citation rate only)	
	Bibliographic description

Research object & Aim:

Research object: Vision transformer with Explainable Artificial Intelligence (XAI) for breast cancer detection to enhance accuracy and interpretability.

Aim: To develop and evaluate a vision transformer-based technique used to improve breast cancer detection, integrated XAI to enhance interpretability. A mathematical model Fisher-KPP a reaction diffusion equation-based technique used to generate spatiotemporal breast tumor growth concentration maps and train an attention U-Net model on it.

Research Tasks:

- Propose a method based on a vision transformer-based approach was used to improve breast cancer diagnostic accuracy and enhance interpretability using XAI. Evaluate the proposed technique using different datasets and compare the obtained results with existing methods used for breast cancer detection.
- A simulation driven learning framework is proposed in which Fisher-KPP model was used to generate the breast tumor spatiotemporal concentration maps based on mammographic images from CBIS-DDSM dataset, tumor location, temporal information, proliferation rate, and diffusion coefficient. The trained modified Attention U-Net model generates breast tumor spatiotemporal concentration using the concentration maps generated from simulation. A visual interpretability analysis was performed using LayerCAM.

Stages of research and dissertation preparation:

Title of the work		Deadlines	Notes
2.2	Theoretical study:	December 2025 - September 2026	
2.2.1	Analysis and systematization of scientific and other information		A comprehensive review and systematization of scientific literature related to breast cancer detection using artificial intelligence, deep learning models, vision transformers, and explainable AI techniques is completed, and findings are published in a review article.
2.2.2	Exploring Breast Cancer Detection Techniques using Explainable AI		Study in depth knowledge of Explainable Artificial Intelligence techniques (Grad-CAM, Grad-CAM++, integrated gradients, SHAP) used in medical imaging to explain the black-box model decision and improve interpretability in breast cancer detection.
2.2.3	Parameterization of Breast Cancer Detection Using Vision Transformers and Explainable AI Techniques.		The parameterization and optimization of vision transformer-based models are studied to improve model performance and reduce computational power. Additionally, Hyperparameter optimization is performed using Optuna, which provides an efficient and computationally effective alternative to traditional grid or random search methods for complex deep neural networks.

Title of the work		Deadlines	Notes
2.3	Empirical study:	January 2026 - June 2027	
2.3.1	Description of the experimental part		The implementation and evaluation of the vision transformer-based model are in progress. Some initial steps, data preprocessing, data augmentation, training, and parameter selection. To assess the performance of the proposed model, we examined it in terms of evaluation metrics.
2.3.2	Realization of the existing method for comparison of the developed method.		To evaluate the performance of the proposed model, the same strategy will be applied to another dataset or existing methods for comparison in terms of evaluation metrics and computational cost used for breast cancer detection.
2.3.3	Creating a method modification.		Based on the initial results from the proposed method, the proposed model will be modified to improve some evaluation metrics and achieve better performance used for breast cancer detection.

Title of the work		Deadlines	Notes
2.3.4	Experimental study of the developed modifications and comparison of the modifications with published existing methods.	January 2026 - June 2027	The proposed model will be compared with an existing method used for breast cancer detection based on evaluation metrics and computational cost. After obtaining all results from the proposed method, the results will be published in a research article, and part of the results is presented at the international conference.
2.4	Analysis of the obtained data, generalization, preparation of conclusions:	July 2026 - June 2027	
2.4.1	Summary of theoretical research.		The theoretical research is mainly focused on two main methodologies: First, the vision transformer-based approach was used to improve diagnostic accuracy including the concepts of SMOTE, token learner, self-attention and model interpretability using Gradient-weighted Class Activation Mapping (Grad-CAM).

Title of the work		Deadlines	Notes
2.4.1		July 2026 - June 2027	Second, a mathematical model Fisher-KPP was used to generate spatiotemporal tumor growth concentrations using biological parameters and CBIS-DDSM dataset.
2.4.2	Summary of empirical research.		In a vision transformer-based approach, a TokenLearner combined with multi-layer perceptron is used to develop an efficient model that requires less computational power compared to traditional vision transformer for breast cancer detection. To make the proposed model interpretable, Grad-CAM was used. A modified Attention U-Net framework was used for predicting future spatiotemporal breast tumor growth using mammograms from the CBIS-DDSM dataset and tumor growth concentration maps generated from Fisher-KPP model. To better understand the proposed modified Attention U-Net framework decision making process of predicted tumor-concentration maps, a visual interpretability analysis was performed using Layer-wise Class Activation Mapping (LayerCAM).

Title of the work		Deadlines	Notes
2.4.3	Summarization of results, preparation of conclusions.	July 2026 - June 2027	In Vison transformer-based approach the model achieved a mean accuracy 0.9774 ± 0.0058, precision 0.9734 ± 0.0101, recall 0.9944 ± 0.0032, F1-score 0.9838 ± 0.0041 and AUC 0.9978 ± 0.0006 for binary classification. A modified attention-Net framework achieved the test Loss 0.20, accuracy: 0.99, Precision:0.78, Recall 0.80, F1-score and Dice coefficient 0.77, IoU 0.73 and MAE 0.001.

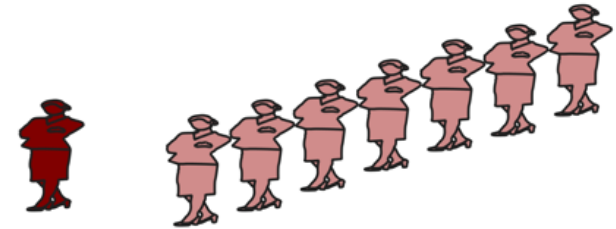
Description of scientific results obtained during the six-month period:

- A vision transformer-based approach was applied to improve breast cancer diagnostic accuracy with less computational resources. A Gaussian-based Synthetic Minority Oversampling Technique (SMOTE) is used to generate synthetic images and balance the BreakHis dataset. A TokenLearner combined with multi-layer perceptron is used to develop an efficient model that requires less computational power compared to traditional vision transformer for breast cancer detection. To make the proposed model interpretable, Grad-CAM was used to show the critical regions where model focuses to extract the most important tokens of the possible prediction outcomes.
- A mathematical breast tumor simulation model based on Fisher-KPP to generate breast tumor concentration maps using mammograms from the CBIS-DDSM (Curated Breast Imaging Subset of the Digital Database for Screening Mammography) dataset, which are then used to train a modified Attention U-Net framework for predicting future spatiotemporal breast tumor growth using mammograms from the CBIS-DDSM dataset and simulation maps. To better understand the proposed modified Attention U-Net framework decision making process of predicted tumor-concentration maps, a visual interpretability analysis was performed using LayerCAM.

Work Plan for the next half year

- To participate in an international conference and present research findings on a mathematical breast tumor simulation model based on Fisher-KPP reaction diffusion equation which was used to generate the breast tumor concentration maps using temporal information, proliferation rate and mammograms from CBIS-DDSM dataset.
- Publish a scientific article in a peer reviewed journal.

According to World Health Organization (WHO) Breast Cancer (BC) is the most prevalent disease worldwide, with over 2.3 million new cases annually. The most significant risk factor for breast cancer is being a woman. Women are affected by breast cancer in about 99% of cases, whereas men are affected in 0.5–1% of cases. Women with breast cancer who live in high-income nations have a 60% higher chance of surviving than those who live in low and middle-income countries [1].



**Breast cancer affects 1 in 8
women**

Artificial intelligence (AI), particularly Deep Learning (DL), demonstrate the potential to enhance the diagnostic and prognostic accuracy to detect BC [1,2]. However, its clinical implementation remains limited due to its "black-box" nature. The Vision Transformer (ViT), a DL model based on a multi-head attention mechanism, has been extensively employed in several computer-aided diagnosis tasks due to its strong capability to effectively process medical imaging data [3,4]. Due to their architectural complexity, ViT relies on powerful GPUs or CPUs to enable fast training and implementation in real-world clinical diagnostic applications.

In this study, proposed model addresses these limitations, delivering high accuracy and interpretability while significantly reducing the required computing power. Our evaluation focuses on key performance metrics, interpretability and computational efficiency. By utilizing interpretable algorithm, this approach supports clinicians in making informed decisions, minimizing diagnostic errors, and improving patient outcomes.

- [1]. Dastgeer, S. and Treigys, P. Llms and xai for breast cancer transparency: areview. *Baltic journal of modern computing.*, 13(2):528–550, 2025.
- [2]. Ahmed, F., Naz, N. S., Khan, S., Rehman, A. U., Ismael, W. M., and Khan,M. A. Explainable artificial intelligence (xai) in medical imaging: a systematic review of techniques, applications, and challenges. *BMC MedicalImaging*, 2026.
- [3]. Dastgeer, S. and Treigys, P. Transforming black-box models into explainableai for breast cancer recognition. In *DAMSS: 15th conference on data analysis methods for software systems*, Druskininkai, Lithuania, November 28-30,2024., pages 19–20. Vilniaus universiteto leidykla, 2024.
- [4]. Aburass, S., Dorgham, O., Al Shaqsi, J., Abu Rumman, M., and Al-Kadi,O. Vision transformers in medical imaging: a comprehensive review of advancements and applications across multiple diseases. *Journal of ImagingInformatics in Medicine*, 38(6):3928–3971, 2025.

Contribution:

This study proposes an interpretable hybrid deep learning model with Gradient-weighted Class Activation Mapping (Grad-CAM) for BC classification using BreakHis dataset. With the integration of Grad-CAM, the proposed model highlights the specific regions within an image that highly contributes to model outcome, providing interpretable insights that support medical professionals to strengthen clinical evaluation.

Results:

The proposed model achieved a mean accuracy 0.9774 ± 0.0058 , precision 0.9734 ± 0.0101 , recall 0.9944 ± 0.0032 , F1-score 0.9838 ± 0.0041 and AUC 0.9978 ± 0.0006 for binary classification.

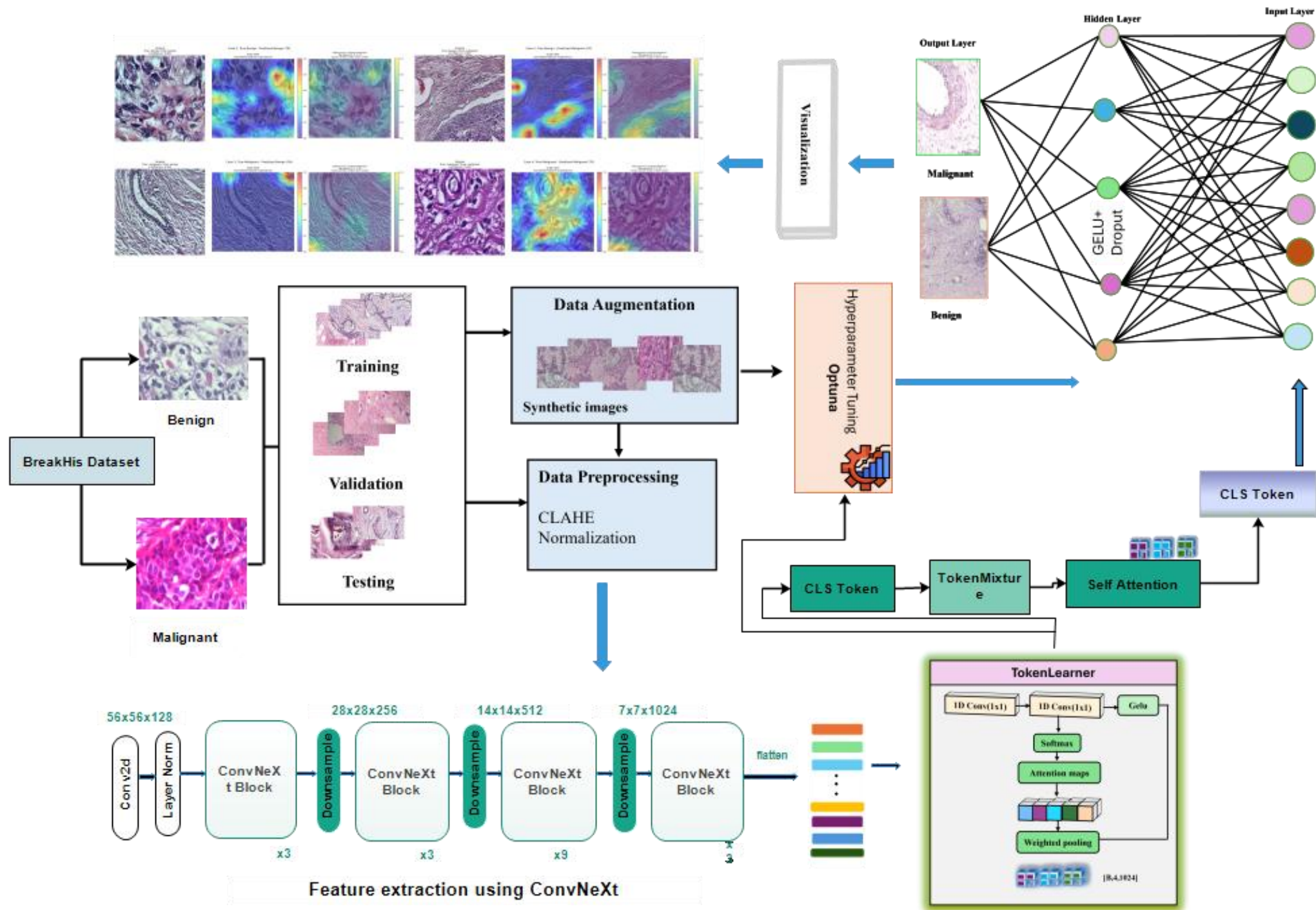
Methodology:

A Gaussian-based Synthetic Minority Oversampling Technique (SMOTE) is applied to generate synthetic minority samples to balance the BreakHis dataset.

Additionally, Contrast Limited Adaptive Histogram Equalization (CLAHE) is employed to enhance image contrast and mitigate the effects of noise and low image quality.

Hyperparameter optimization is performed using Optuna, which provides an efficient and computationally effective alternative to traditional grid or random search methods for complex deep neural networks. The optimization included the training, sampling and architectural parameters such as learning rate, tokens, number of MLP layers, MLP hidden dimension, attention head and m which were initialized as 0.00005, 4, 2, 256, 8 and 3 respectively.

The second step involves feature extraction, where ConvNeXt is used to extract high-level histopathological features from BreakHis dataset. A TokenLearner module is used to identify and extract the most important tokens based on the attention mechanism. Subsequently, a learnable classify token (CLS) is added to sequence of a learned tokens to capture the global contextual information. The entire set of tokens is then passed through a TokenMixer module based on multi-head self-attention, and the refined CLS token embedding is fed into a multi-layer perceptron (MLP) classifier to categorize breast tumor cells as benign or malignant.



Conclusion

This research demonstrated the effectiveness of an interpretable and efficient framework with improved accuracy for breast cancer classification using microscopic images. By integrating data preprocessing, TokenLearner and a ViT-based classification, the proposed model enhances interpretability and clinical applicability for breast cancer diagnosis. Future studies should aim to validate the model on more complex and diverse datasets.



Thank you.

