ViT-ESA: Enhanced Spatial Attention in ViT for Breast Cancer Histopathological Image Classification

Amrutanshu Panigrahi[1], Subrat Chowdhury[2]
[1]Postdoctoral Researcher, Lincoln University College, Malaysia; [2]Sri Venkateswara College Of Engineering & Technology (Autonomous), Chittoor, AP, India
Email ID: amrutansup89@gmail.com, subrata895@gmail.com

Abstract: Breast cancer is one of the most common causes of death due to cancer among women in the world requiring accurate and early diagnosis. Accurate diagnosis of histopathological images is crucial for final diagnosis. Moreover, many computer diagnosis systems do not capture contextual information globally or at several scales of tissue pattern. Recent studies showed that ViT's models which use deep learning show a better performance in modeling long-range dependencies than CNN's. Even though high-dimensional deep features can improve performance through improved feature representation, they also introduce redundancy and computation overloads. These can hinder clinical deployment. To address these concerns, the proposed method integrates deep feature extraction based on Vision Transformers with feature selection using the Elephant Search Algorithm (ESA). Throughout the BreakHis breast histopathology dataset when tested at various magnification factors, the proposed ViT–ESA framework improves classification performance at lower feature dimensions. A comparative evaluation of multiple machine learning classifiers demonstrates that the XGBoost classifier performs the best validating the effectiveness of this method with ESA-based transformer features for reliable breast cancer diagnosis.

Keywords: Breast cancer; Vision Transformer; Elephant Search Algorithm; Feature optimization

Introduction

The histopathological diagnosis of breast cancer image has become an important and at the same time, a challenging act within the medical imaging community due to high intra-class variability, high inter-class similarity, and scale-dependent tissue patterns [1, 2]. Pathologists' manual checks take time and are subject to bias. Intelligent diagnostic and therapeutic systems, based on diagnostic protocols. As far as I know, machine learning (ML) and deep learning (DL) techniques have proven useful for clinicians, assisting them with greater diagnostic accuracy and consistency. Convolutional neural networks are often the most popular choice for histopathological image analysis. However, convolutional neural networks have limited scope of learning global context due to their local receptive fields. In recent years, Vision Transformers have emerged as powerful alternatives that exploit attention to model long-range dependencies across image regions [3]. The use of ViTs is certainly beneficial in practice, but the high-dimensional feature maps produced by such models may contain redundancy or noise. This affects the overall speed of computation as well as generalization [4]. In this study, we address the drawbacks found in the existing literature by proposing an innovative hybrid framework which integrates ViT-based feature extraction with bio–inspired metaheuristic optimization technique, the Elephant Search Algorithm (ESA), for selecting a discriminative feature subset [5, 6].

Motivation

The research work was inspired by the key observations. Histopathological images contain complex global structures an fine-grained texture patterns which cannot be recognized by CNNs alone. ViTs help suppress noise in the background while improving the signal quality from the region of interest in histopathology

images. The use of high dimensional deep features incurs high computational cost and is prone to overfitting with unbalanced and small datasets. Additionally, traditional methods for feature selection and hyperparameter optimization (like grid search and random search) require substantial computational power and risk becoming trapped in local optima. The effectiveness of the Elephant Search algorithm is enhanced by exploring the most promising areas around a point.

Objective

The primary objectives of this research are as follows:

1. To implement the ViT for extracting the features from the Breast Cancer Histopathological images.
2. To apply the Elephant Search Algorithm for optimal feature selection, reducing redundancy and irrelevant information.
3. To improve classification accuracy, robustness, and computational efficiency using ESA-optimized features.
4. To evaluate the effectiveness of the proposed framework using multiple machine learning classifiers and performance metrics.

Methodology

The proposed methodology consists of a multi-stage pipeline comprising data preparation, deep feature extraction with ViT, feature optimization with ESA, and final classification with traditional machine learning classifiers. The overall workflow is designed to ensure robustness, scalability, and reproducibility. Table 1 shows the dataset description.

Table 1. Dataset Description

Magnification	Benign	Malignant
40×	625	1370
100×	644	1437
200×	623	1390
400×	588	1232
Total	**2480**	**5429**

Vision Transformer (ViT)

A vision transformer is used as a deep feature extractor by splitting the image into fixed-size patches without overlap. After linearly embedding each patch and summing with positional encodings, we feed them into multiple transformer encoder layers. The ViT self-attention mechanism captures long-range dependencies and global context between image patches. A pretrained ViT model is used in this study without the classification head. Feature embeddings are extracted from the final transformer encoder layer to create high-dimensional feature vectors that capture the full characteristics of histopathological tissue structures. The features act as input for the next optimization phase [7].

Elephant Search Algorithm (ESA)

The elephant herding behaviour is the basis of the elephant herding optimization. The elephants live in clans led by a matriarch who is responsible for exploitation (local search) which is the main reason why males promote exploration (global search). This equalization permits an effective convergence and minimizes local minima. In the proposed framework, the ESA is employed for the selection of features. Each elephant corresponds to a binary vector signifying which features were chosen and which were deleted. Classification performance is maximized and feature dimensionality is minimized. By utilizing clan updating and class-separating operations multiple times in the class-finding algorithm, ESA can select a better feature subset that reduces computation and increases generalization [8].

Results and Discussion

The optimized feature sets using ESA are evaluated on a number of classifiers, including Support Vector Machine (SVM), Random Forest (RF), AdaBoost, Gradient Boosting, XGBoost, and Extreme Learning Machine (ELM). Evaluation is done on data using accuracy, balanced accuracy, precision, recall, F1-score, ROC curve, and precision–recall curve. . Figure 1 shows the confusion matrix. Figures 2 and 3 show the ROC and PRC analysis of the proposed model. Figure 4 shows the ESA convergence over different iterations. Figure 5 shows the feature importance of selected features from ESA.

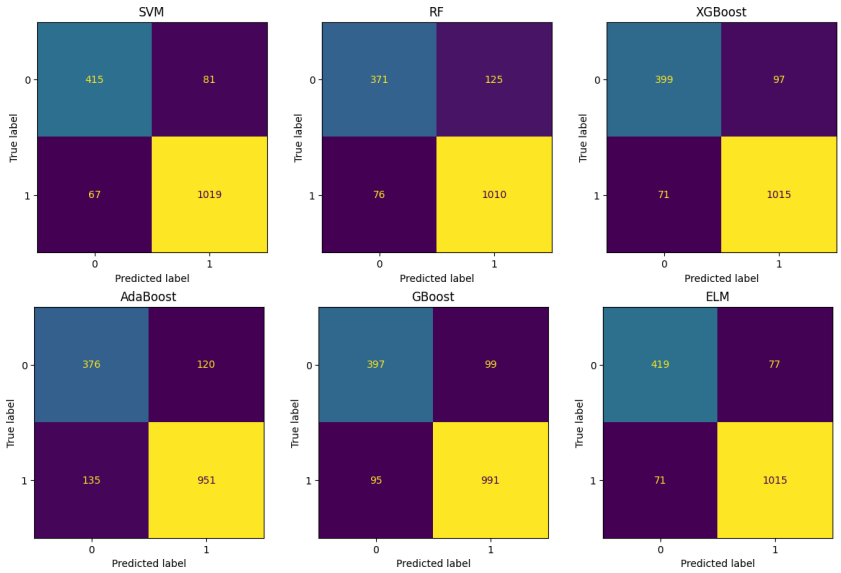

Figure 1. Confusion matrix for different classifiers

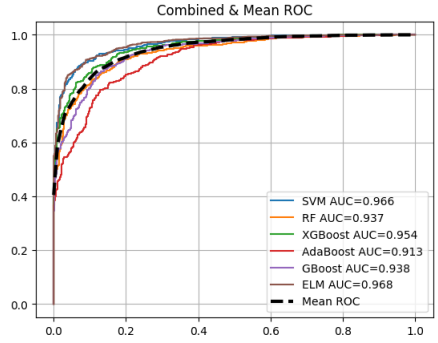

Figure 2. ROC of the proposed model

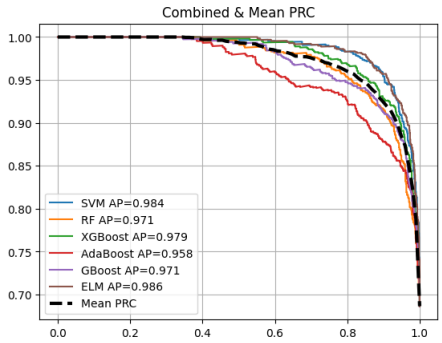

Figure 3. PRC analysis of the proposed model

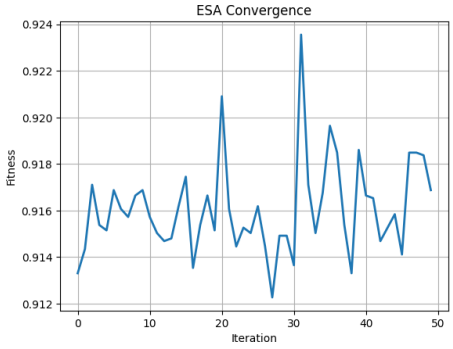

Figure 4. ESA Convergence for 5iterations

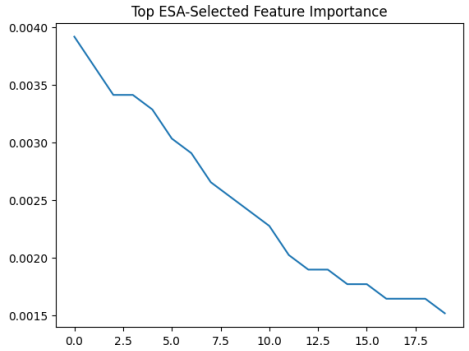

Figure 5. Top ESA-Selected Feature Importance

Conclusions

This paper proposes a new robust and hybrid framework using Vision Transform (ViT) and Elephant Search Algorithm (ESA) for breast cancer histopathology image classification. The study presents a multi-scale global convolutional neural network-based deep learning model (CNN-BDL) capable of generating output with a resolution of 512 pixels greater than its input, a 256x256 pixel U-Net CT image. Based on experimental results conducted on the BreakHis dataset, it is evident that the ESA-optimized ViT features considerably increase the classification performance of a range of classifiers, including a variety of machine learning classifiers. The XGBoost results were better than all others, and the balanced accuracy values indicate that the class imbalance was handled effectively. The suggested framework demonstrates significant deployment capability in a real clinical environment for intelligent decision-support systems in sustainable healthcare. Future work will include incorporating explainable AI (XAI) techniques, integrating multimodal data, and conducting large-scale clinical validation to further improve reliability and interpretability.

References

1. Panigrahi, A., & Chowdhury, S. (2025). Machine Learning and Deep Learning Approaches for Breast Cancer Diagnostics: A Systematic Review. SGS-Engineering & Sciences, 1(3).
2. Hayat, M., Ahmad, N., Nasir, A., & Tariq, Z. A. (2024). Hybrid deep learning EfficientNetV2 and vision transformer (EffNetV2-ViT) model for breast cancer histopathological image classification. IEEE Access, 12, 184119-184131.
3. Yan, Y., Lu, R., Sun, J., Zhang, J., & Zhang, Q. (2025). Breast cancer histopathology image classification using transformer with discrete wavelet transform. Medical Engineering & Physics, 138(1), 104317.
4. Ogundokun, R., Owolawi, P., & Tu, C. (2025). Optimized deep feature learning with hybrid ensemble soft voting for early breast cancer histopathological image classification. Computers, Materials, & Continua, 84(3), 4869.
5. Ejiyi, C. J., Cai, D., Fiasam, D. L., Adjei-Arthur, B., Obiora, S., Ayekai, B. J., ... & Qin, Z. (2025). Multi-modality medical image classification with ResoMergeNet for cataract, lung cancer, and breast cancer diagnosis. Computers in Biology and Medicine, 187, 109791.
6. Yusuf, M., Kana, A. F. D., Bagiwa, M. A., & Abdullahi, M. (2025). Multi-classification of breast cancer histopathological image using enhanced shallow convolutional neural network. Journal of Engineering and Applied Science, 72(1), 24.
7. Othman, E. M. (2024). Breast Cancer Multi-Class Classification Using ViT Model. International Journal of Computer Applications, 186(13), 13-18.
8. Tian, Z., Fong, S., Wong, R., & Millham, R. (2016, August). Elephant search algorithm on data clustering. In 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 787-793). IEEE.