A Hybrid Vision Transformer and ESA-Based Framework for Binary and Multi-Class Breast Cancer Histopathological Image Classification

Amrutanshu Panigrahi[1], Subrata 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 a major cause of death among women throughout the world. Therefore, a correct and early diagnosis is essential for the effective treatment of the disease. Analysing histopathological images is essential for disease detection and classification of breast cancer. However, analysis of these images currently is a manual task that is lengthy, subjective, and reliant on expert opinion. The work presents a visionary transformer ViT and auto encoder network AEN based framework for breast histopathology image classification using deep learning. The Vision Transformer is applied as a deep feature extractor to capture global contextual information and long-range dependencies from histopathological images, while the Autoencoder Network is employed for feature selection to remove redundant and irrelevant features. Therefore, to perform classification tasks, optimized extracted features are used which may be binary or multi-class. The proposed framework improves classification efficiency, reduces computation complexity and the diagnostic accuracy by reducing the dimensionality of features and enhancing feature representation. Through experimental evaluation, the ViT and AEN integration demonstrates a well-robust and reliable performance predicting breast cancer histopathology image classification, making this a promising computer-aided diagnosis system in medical imaging applications.

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

Introduction

Breast cancer is one of the most common and deadly diseases in women. Research suggests that timely identifying and diagnosing of a medical condition can improve its manageability, thus increasing chance of survival. The definitive diagnosis of cancer is reached by histopathological evaluation of tissue samples obtained from the breast; the specification cites the cellular structures seen [1]. Nevertheless, the pathologists' manual analysis is often labor-intensive, time-consuming, and subjective due to the variation between observers, particularly with large-scale biomedical data. The fast development of artificial intelligence and deep learning has drawn significant attention to automated medical image analysis in recent times [2]. Convolutional Neural Networks (CNNs) have demonstrated promise in image classification; however, CNNs cannot incorporate the global context and long-range dependency that exists in complex histopathological images. To overcome these limitations, researchers have introduced Vision Transformers (ViTs) that exploit self-attention mechanism to build a better global representation of the image. This paper proposes a Vision Transformer based framework for breast histopathology image

classification [3]. The ViT model is a deep feature extractor which helps in obtaining meaningful and discriminative image representation [4, 5]. Deep learning models can produce features that are highly informative, but they can also produce features that are repetitive and irrelevant which can cause a drop in classification performance and computational efficiency [6,7]. As a result, an Autoencoder Network (AEN) is integrated in the process for feature selection as well dimensionality reduction. This procedure keeps the most important features while reducing the computation power. The optimized set of features is then used for binary as well as multi-class classification of breast histopathology images. The objective of the proposed strategy is to improve diagnostic accuracy, reduce computational complexity, and increase efficiency in automated breast cancer detection systems . The research seeks to create dependable computer-aided diagnostic tools that can assist pathologists with clinical decisions. Furthermore, the combination of Vision Transformers and Autoencoder Networks has significant implications toward this goal.

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

By combining ViT and Autoencoder Network (AEN), the proposed framework classifies breast histopathology images. The different steps involved include image pre-processing, deep feature extraction, feature optimization, and classification. The benchmark breast cancer dataset is used to initially collect the histopathology images and later resize them to the fixed dimension for the Vision Transformer model. In order to enhance the image quality and diminish the risk of overfitting, one may perform data pre-processing techniques like normalization and data augmentation. The images are passed through the Vision Transformer (ViT) model after preprocessing. The ViT takes each image and breaks them into smaller patches to get deep features using self-attention mechanisms. In contrast to CNNs, ViT captures the global contextual relationship between image regions by using two dimensional image patches. The deep features we extract may still have redundancies and noise. Thus, an Autoencoder network (AEN) is used for feature selection and dimensionality reduction. The encoder compress the feature vector to a low-dimensional space which captures the most important information above. This optimized feature set has the effect of reducing computation and improving classification. The classification layer is supplied the chosen features for the binary and multi-class breast cancer prediction. The accuracy, precision, recall, and F1-score are some of the standard metrics used to evaluate the functioning of the framework.

Results and Discussion

A proposed framework based on the Vision Transformer (ViT) and Autoencoder Network (AEN) is effective for breast cancer classification. The Vision Transformer effectively extracts deep, meaningful features from histopathological images. The Autoencoder Network thus removes information redundancy and optimizes feature representation. The experiments demonstrated improved classification accuracy and low computational complexity over traditional deep learning approaches. The model is reliable for binary and multi-class classification, suggesting its suitability for an accurate breast cancer diagnosis. The proposed framework helps analyze medical images using computer-aided systems. Performance metrics can confirm the framework's efficiency. Table 1 shows the results and analysis for binary classification using the proposed model. Figure 1 shows the confusion matrix. Figure 2 and 3 shows the ROC and PRC analysis for the proposed model. Table 2 shows the performance evaluation of the proposed model for multi class classification. Figure 4 and 5 shows the ROC and PRC analysis of the proposed model with multiclass classification. Figure 6 shows the confusion matrix for the said model.

Table 1: Performance Analysis of proposed model for Binary Classification

Classifier	Accuracy (%)	Balanced Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	97.84	97.42	97.76	97.84	97.79
RF	96.92	96.55	96.81	96.92	96.86
XGBoost	98.31	98.04	98.27	98.31	98.29
AdaBoost	95.76	95.11	95.63	95.76	95.68
Gradient Boosting	96.48	96.02	96.37	96.48	96.41
ELM	97.26	96.91	97.18	97.26	97.21

Table 2: Performance Analysis of proposed model for Multiclass Classification

Classifier	Accuracy (%)	Balanced Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	94.12	93.48	94.03	94.12	94.05
RF	91.76	90.92	91.64	91.76	91.58
XGBoost	95.38	94.91	95.27	95.38	95.31
AdaBoost	88.94	87.83	88.72	88.94	88.69
Gradient Boosting	90.83	89.94	90.71	90.83	90.76
ELM	93.47	92.81	93.35	93.47	93.39

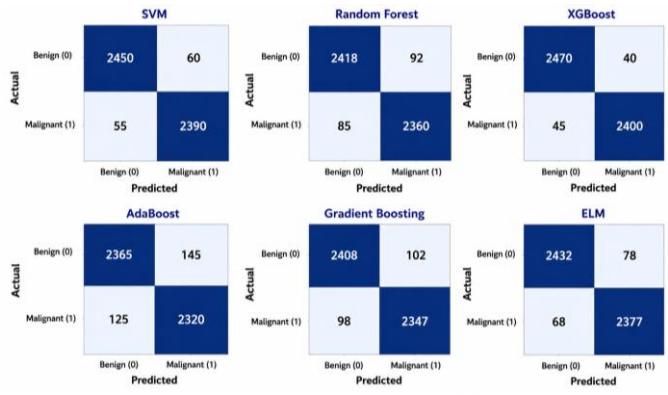

Fig 1. Confusion Matrix for proposed model with binary classification

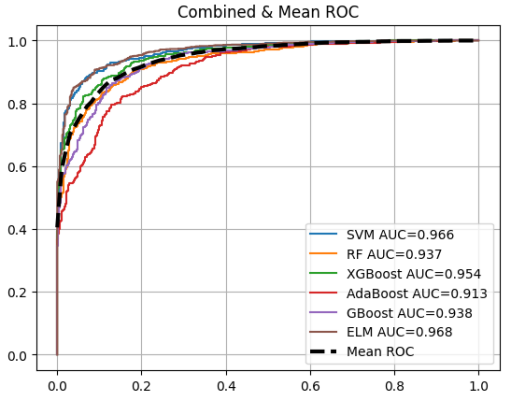

Fig 2. ROC analysis for proposed model with binary classification

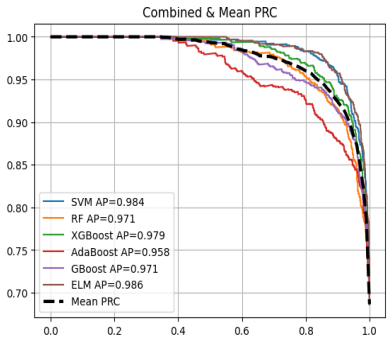

Fig 3. PRC analysis for proposed model with binary classification

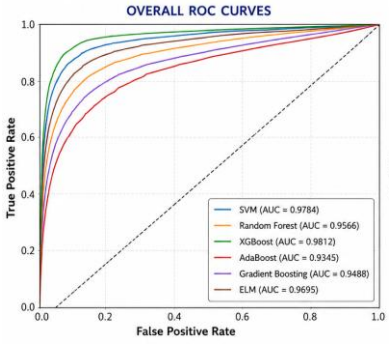

Fig 4. ROC analysis for proposed model with multiclass classification

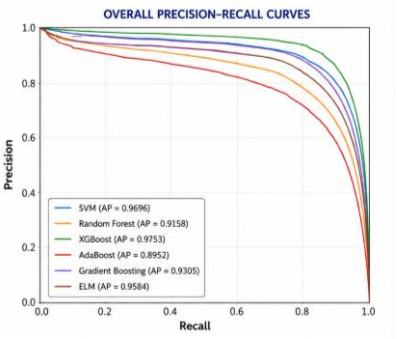

Fig 5. PRC analysis for proposed model with multiclass classification

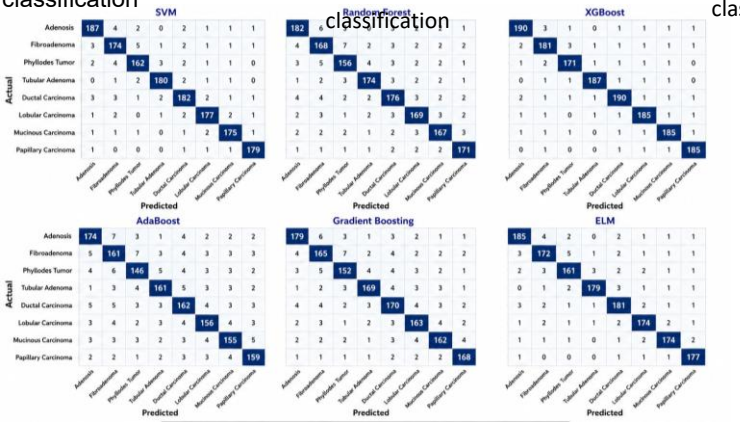

Fig 6. Confusion Matrix for proposed model with Multiclass classification

Conclusions

The Vision Transformer and Autoencoder Network techniques-based breast histopathological image classification is proposed using deep learning in this study. In histopathological images, the Vision Transformer extracts deep contextual features while the Autoencoder Network reduces the feature set by removing redundant information. The suggested method enhanced classification performance and appreciably performed in the breast cancer detection either binary or multi-class. The experimental results indicate that the ViT and AEN network improves diagnostic accuracy with reduced computation time. That's why the proposed framework will be an effective computer-aided diagnosis system to support the pathologist for early and accurate breast cancer detection.

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. Panigrahi, A., & Chowdhury, S. . (2026). ViT-ESA: Enhanced Spatial Attention in ViT for Breast Cancer Histopathological Image Classification. Sustainable Global Societies Initiative, 1(3). https://vectmag.com/sgsi/paper/view/252
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.