Anomaly Detection in Breast Histopathology Images with Convolutional Variational Autoencoders

Diptanshu Sikdar, Travis Tran, James Xu, Jordan Yee
Department of Computer Science, University of California, Irvine

Abstract

We explore Convolutional Variational Autoencoders (ConvVAEs) for anomaly detection in breast histopathological images, comparing their performance against Fully Connected VAEs (FC-VAEs). We implement four primary models: (1) a baseline Variational Autoencoder (VAE), lacking spatial awareness; (2) a standard ConvVAE, leveraging convolutional layers for feature extraction; (3) a VAE with a Pre-trained U-Net Encoder, utilizing a pretrained ResNet34 encoder for improved representation learning; and (4) an Attention-enhanced ConvVAE, integrating attention mechanisms to enhance feature learning. Each model encodes image data into a latent space, allowing for reconstruction-based anomaly detection. Using a publicly available histopathology dataset, we compare these architectures by evaluating their reconstruction quality and ability to distinguish cancerous from non-cancerous samples. Our preliminary results indicate that the vanilla and attention-based ConvVAEs significantly outperform the traditional VAE, confirming the necessity of convolutional structures for effective feature extraction. However, among the convolution-based models, no single architecture demonstrates a consistently superior performance over all metrics.

Introduction

Breast cancer is the **most commonly diagnosed cancer among women worldwide**, accounting for approximately 2.3 million new cases and nearly 685,000 deaths in 2020 alone. **Early detection significantly improves survival**, yet the current standard, which is manual histopathological evaluation, is subjective and labor-intensive. Deep learning offers promising solutions, but traditional supervised models require large, expert-annotated datasets, which are scarce. Unsupervised approaches like Variational Autoencoders (VAEs) overcome this by detecting anomalies as deviations from learned patterns of normal tissue. However, standard VAEs lack the ability to capture complex spatial structures in histopathology images. To address this issue, we utilized **Convolutional Variational Autoencoders (CVAEs)**, which can capture essential spatial features for improved breast cancer diagnostics. We explored several kinds of CVAE architectures:

- Vanilla CVAE with Conv2D and BatchNorm2D layers
- Custom CVAE with a pre-trained U-Net encoder
- Custom CVAE with Attention layers after each convolutional layer

Our study evaluates these approaches on a publicly available breast histopathology dataset. The results demonstrate that incorporating reconstruction loss with Kullback-Leibler divergence (KL divergence) significantly enhances anomaly detection performance. All ConvVAE models except the ones with U-Net encoders achieved superior performance metrics compared to the baseline VAE in terms of F1 and AUC scores.

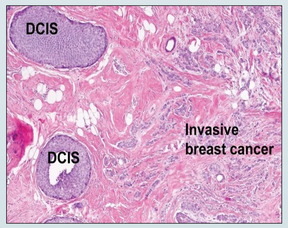


Figure 1: Breast tissue histopathology sample with Ductal Carcinoma In Situ (DCIS)

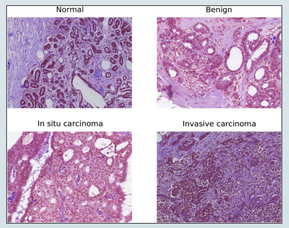


Figure 2: Four breast pathology images at different stages

Methodology

A. Variational Autoencoder (VAE)

As a baseline model, we implemented a traditional Variational Autoencoder (VAE) to assess the impact of convolutional architectures on feature extraction and anomaly detection. Unlike convolutional VAEs, which leverage spatial feature hierarchies, this model consists entirely of fully connected layers, treating images as flattened vectors.

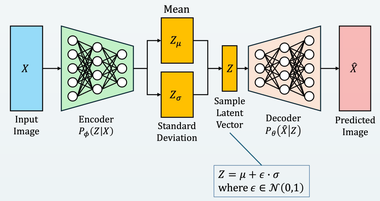


Figure 3: Traditional Autoencoder Architecture

B. Vanilla Convolutional VAE (ConvVAE)

The ConvVAE encoder extracts hierarchical image features using four convolutional layers with Batch Normalization and ReLU activations, progressively reducing spatial dimensions via strided convolutions. The final output is flattened and passed through two fully connected layers to generate the mean (μ) and log variance (σ^2) parameters of the latent distribution. The latent space is sampled using the reparameterization trick, which enables gradient-based optimization.

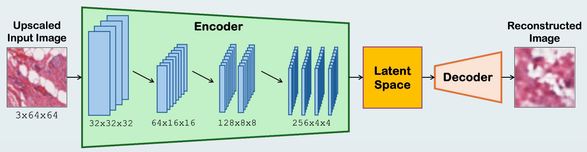


Figure 4: Vanilla Convolutional VAE Architecture

The decoder reconstructs input images from the latent space using four transposed convolutional layers with Batch Normalization and ReLU activations, followed by a final Tanh activation to produce images in the range [-1, 1]. Unlike standard VAEs, the ConvVAE utilizes convolutional layers to effectively capture local and hierarchical spatial features essential for distinguishing cancerous histological patterns.

C. VAE with a U-Net Encoder (ConvVAE-U-NET)

In this model, a pretrained U-Net with a ResNet-34 backbone is integrated into the encoder of the baseline ConvVAE model. Unlike conventional encoders, this architecture can potentially capture multi-scale, hierarchical spatial features, benefiting from earlier knowledge of cell structures. The ConvVAE-U-NET consists of three components: a pretrained ResNet-34 feature extractor with an optional weight-freezing mechanism, a fully connected latent transformation, and a simple decoder, as shown below.

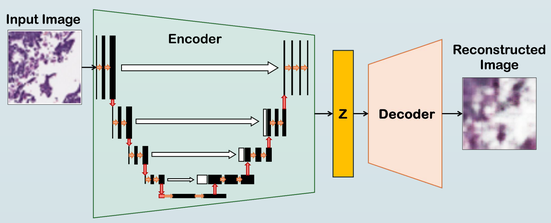


Figure 5: ConvVAE with Pre-trained U-NET Encoder

D. Attention-Enhanced Convolutional VAE (Attn-ConvVAE)

Building upon the ConvVAE, we introduce an attention mechanism to enhance spatial feature extraction. Specifically, self-attention layers are placed after each convolutional layer to allow the model to learn important local features on a hierarchical basis.

Experiments

A. Dataset

We used the Breast Histopathology Images dataset from Kaggle, containing 277,524 image patches (50x50 pixels) labeled by either the presence or absence of **invasive ductal carcinoma (IDC)**. Since our focus is anomaly detection, models were trained only on non-cancerous images; thus, higher reconstruction losses during testing indicated potential cancerous anomalies.

B. Experimental Setup

The data is split into 80% training, 10% validation, and 10% test, and all 50×50 images are upscaled to 64×64 using **bicubic interpolation** to improve image quality and compatibility with convolutional architectures. Models are trained using the Adam optimizer with a learning rate of 1×10^{-5} and evaluated using the **Wasserstein Earth Mover's Distance** to identify anomalies. We determine the optimal anomaly threshold using **Youden's J statistic**, **balancing sensitivity (TPR)**, and **specificity (FPR)**.

C. Evaluation Metrics

We assess our model's ability to distinguish normal and anomalous data using several metrics:

- **Reconstruction Loss (MSE)**: Measures how accurately images are reconstructed. Lower values indicate better reconstruction for normal data; higher values help identify anomalies.
- **Kullback-Leibler Divergence (KL Divergence)**: Regularizes the latent space to match a prior distribution, preventing overfitting and aiding generalization.
- **Confusion Matrix**: Counts True Positives, True Negatives, False Positives, and False Negatives to analyze model biases and classification errors.
- **F1 Score**: Balances precision and recall, essential for imbalanced medical datasets.
- **Area Under the Curve (AUC)**: Evaluates the model's ability to discriminate between normal and anomalous samples across classification thresholds.
- **Accuracy**: Reflects the proportion of correctly classified samples but is interpreted alongside other metrics due to potential class imbalance.

Results

Here are the results of our models on the Breast Histopathology Dataset.

Table I: Model Comparisons

Model	VAE	ConvVAE	Attn-ConvVAE	Frozen CVAE-U-Net	Unfrozen CVAE-U-Net
Reconstruction Loss	1075.94	424.14	374.57	639.11	540.59
KL Divergence	48.71	261.09	392.12	20.01	12.35
Accuracy (%)	54.43	65.58	57.89	51.78	53.99
F1 Score	0.4125	0.6431	0.6328	0.6533	0.6395
AUC	0.53	0.70	0.60	0.50	0.53

A. Reconstruction Loss

Attn-ConvVAE (374.57) has the lowest reconstruction loss, followed by ConvVAE (424.14), both significantly outperforming the baseline VAE (1075.94). U-Net models show moderate performance.

B. KL Divergence

U-Net models achieve the lowest KL divergence (Frozen: 20.01, Unfrozen: 12.35), effectively regularizing the latent space. ConvVAE (261.09) and Attn-ConvVAE (392.12) prioritize reconstruction over regularization.

C. Accuracy

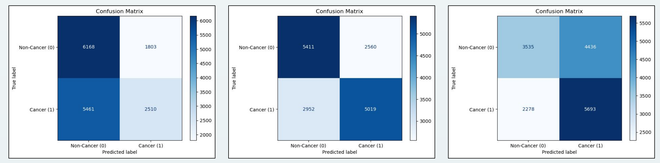
ConvVAE has the highest accuracy (65.58%), outperforming Attn-ConvVAE (57.89%), baseline VAE (54.43%), and U-Net models (~52-54%). Attention provides limited benefits.

D. F1 Score

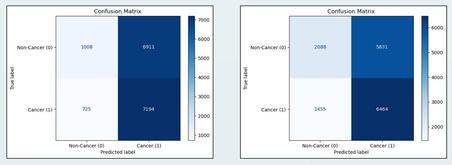
Frozen CVAE-U-Net achieves the highest F1 (0.6533), closely followed by ConvVAE (0.6431), Unfrozen CVAE-U-Net (0.6395), and Attn-ConvVAE (0.6328). VAE (0.4125) lags behind.

E. AUC

ConvVAE demonstrates the highest AUC (0.70), indicating strong anomaly detection. Attn-ConvVAE (0.60) and baseline VAE (0.53) show moderate performance, while U-Net models perform lowest (0.50-0.53).



(a) VAE (b) Vanilla ConvVAE (c) Attn-ConvVAE



(d) Frozen CVAE-U-Net (e) Unfrozen CVAE-U-Net

F. Confusion Matrices

ConvVAE shows strong class separation and fewer errors compared to baseline VAE which struggles with false negatives. Attn-ConvVAE has moderate false positives. U-Net models frequently predict cancerous images resulting in poor discrimination. ConvVAE performs best overall but Attn-ConvVAE has fewer false negatives.

Conclusion

In this study, we explored Convolutional Variational Autoencoders (ConvVAEs), Attention-Enhanced ConvVAEs, and U-Net-based models for anomaly detection in breast histopathology images. ConvVAEs significantly outperformed the baseline VAE, achieving the highest accuracy (65.58%), AUC (0.70), and strong reconstruction capability. The Attention-Enhanced ConvVAE excelled at reconstruction but showed limited improvement in classification. U-Net models effectively regularized latent spaces but struggled with classification accuracy. ConvVAE provided the best balance overall. Future research should focus on higher-resolution images, enhanced attention mechanisms, and input transformations such as denoising and contrast enhancement to further improve model performance.

References

[1] https://doi.org/10.1016/j.breast.2022.08.010
[2] https://doi.org/10.3322/caac.21660
[3] https://doi.org/10.1001/jama.2015.1405
[4] https://doi.org/10.48550/arXiv.1505.04597
[5] https://doi.org/10.3390/app10186427
[6] https://www.tensorflow.org/tutorials/generative/cvae/
[7] https://doi.org/10.3390/ijerph20054244
[8] https://doi.org/10.1038/s41598-020-80610-9
[9] https://doi.org/10.1016/j.comnet.2007.02.001
[10] https://doi.org/10.1109/ICDM.2018.00088
[11] https://doi.org/10.1038/s41598-023-29521-z
[12] http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf
[13] https://doi.org/10.1109/CVPR42600.2020.00813