Anomaly Detection in Breast Histopathology Images with Convolutional Variational Autoencoders

1st Diptanshu Sikdar 2nd Travis Tran 3rd James Xu 4th Jordan Yee
University of California, Irvine, Irvine, USA
dsikdar@uci.edu, travitt1@uci.edu, xujg@uci.edu, jordady1@uci.edu

***Abstract*—We explore the use of Convolutional Variational Autoencoders (ConvVAE or CVAE) for anomaly detection in breast histopathological images, comparing their performance against Fully Connected VAEs (FC-VAEs) and attention-based architectures. We implement four primary models: (1) a baseline Variational Autoencoder (VAE), which lacks 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 Attentive 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 ConvVAEs significantly outperform the simple VAE, confirming the necessity of convolutional structures for effective feature extraction. However, among convolution-based models, no single architecture demonstrates a consistently superior performance across all metrics. These findings underscore the importance of selecting model complexity based on computational constraints and dataset characteristics.**

***Index Terms*—Cancer detection, Convolutional Variational Autoencoder, Anomaly Detection, U-Net, Attention, Histopathology**

I. 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 [1]. Early detection is critical, as the five-year survival rate exceeds 90% when diagnosed at localized stages but drops significantly in later stages. Histopathological examination remains the gold standard for diagnosis, where pathologists analyze stained tissue samples under a microscope to distinguish between benign and malignant cells. However, manual evaluation is time-intensive, subjective, and prone to inter-observer variability, particularly in distinguishing subtle morphological changes [2]. The increasing prevalence of breast cancer underscores the urgent need for automated, scalable, and highly accurate diagnostic tools to assist pathologists in improving diagnostic efficiency and reducing errors.

Deep learning has shown promise in medical image analysis, particularly in histopathology-based cancer detection. However, traditional supervised deep learning models rely on large, high-quality annotated datasets, which are often limited due to the expertise required for labeling. This constraint has led to the exploration of unsupervised learning approaches, such as Variational Autoencoders (VAEs), which learn the underlying distribution of normal tissue and detect anomalies as deviations from expected patterns. While VAEs provide a useful framework for anomaly detection, their fully connected architectures struggle to capture the complex spatial structure of histopathology images. Convolutional Variational Autoencoders (CVAEs) address this limitation by integrating convolutional layers into the encoder and decoder, capturing the essential spatial features embedded in the images.

This study builds upon these advancements by systematically evaluating different CVAE architectures tailored for breast cancer histopathology images. By leveraging convolutional and attention-based mechanisms, we aim to enhance anomaly detection performance, making deep learning-based diagnostics more robust and clinically useful. We explore several different kinds of Convolutional Variational Autoencoder architectures for autonomous anomaly detection in medical imagery:

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

This study evaluates these approaches on a publicly available breast histopathology dataset. Our results demonstrate that incorporating reconstruction loss with Kullback-Leibler divergence (KL divergence) significantly enhances anomaly detection performance. All the CVAE models achieve superior performance metrics compared to the baseline VAE in terms of F1 and AUC scores. Convolutional and minimalistic architectures differ in their capacity to capture clinically relevant anomalies, reinforcing the advantages of CVAEs in medical imaging analysis.

II. Related Works

Anomaly detection (AD) is a topic that, over the last few decades, has become increasingly important in our modern world. The wide variety of applications, from traffic patterns to cybersecurity, have the potential to save lives. At its core, AD is the task of identifying data points that deviate from normality. AD methods usually fall into one of three categories of detection: statistical methods, machine learning methods, and deep learning methods. Here, we give a brief overview of the research and approaches that have already been done.

In concept, statistical methods in anomaly detection maintain two profiles: a stored profile (representing normal behav-

ior) and a current profile (updated with incoming data). An anomaly score is calculated by comparing these two profiles, and an alert is triggered if the differences exceed a predefined threshold. The threshold is typically derived from a statistical distribution model of normal behavior. Some methods assume a Gaussian distribution, using Z-scores or probability density functions to flag anomalies, while others use histogram-based frequency distributions, detecting low-frequency events as outliers. However, statistical methods face limitations. They rely on well-defined distributions, but real-world behavior is often too complex to model accurately. Additionally, these methods often assume static normal behavior, which makes them less effective in dynamic environments. [9]

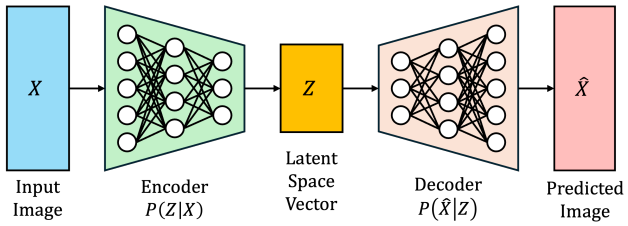


Figure 1: Traditional Autoencoder Architecture

Machine-learning methods were designed to address these concerns. Instead of relying on statistically computed thresholds, anomalies are often detected by the assumption that anomalous data will be isolated from the majority of other data points. Cluster-based methods rely on identifying regions or dense ”clusters” of data points, and calculating distance each data point is to its nearest cluster. This distance is then compared to threshold to determine if that data point is an anomaly. Another method, known as K-Nearest neighbors, records the distances between a data point and neighboring points, judging data points with large k-neared neighbor distances to be anomalies.

More recent approaches, such as those with auto-encoders, fall under the category of deep-learning methods. The goal of these works is to train a model to reconstruct normal data and identify anomalies as reconstructed data with high reconstruction error. [10] Generative Adversarial Networks (GANs) have become more common and are used to model the complex, high-dimensional distributions of real-world data. The GAN model is composed of two competing neural networks, one known as the Generator , and one known as the Discriminator. The Generator attempts to generate real-appearing data, while the Discriminator attempts to discern if that data is real or not. Over time, as the Discriminator improves its detection rate, the Generator will also produce increasingly realistic results. While standard GANs have been successful in modeling natural images and anomaly detection, some methods require solving an optimization problem at test time, making them inefficient for large datasets. Recent approaches, such as the StyleGAN improve the original design through architectural enhancements like better generator normalization, path length regularization, and larger models, leading to higher-quality, more controllable, and more easily invertible image generation. [13]

III. Methodology

A. Goal

The goal of this project is to leverage Autoencoders to learn an accurate representation of histopathology images in a latent space for better anomaly detection and reconstruction. To better adapt to the visual input, the typical architecture of an Autoencoder is altered with layers typically found in Convolutional Neural Networks.

B. 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.

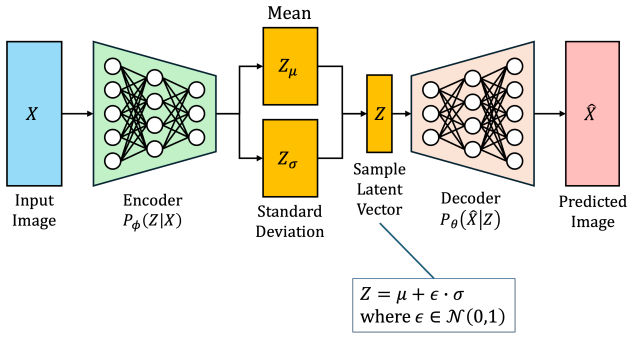


Figure 2: General Variational Autoencoder Architecture

C. Vanilla Convolutional VAE

We implemented the **Convolutional Variational Autoencoder (ConvVAE)** with custom convolutional layers. The model is designed to be able to learn the compact latent representation of histopathology breast tissue images and ideally reconstruct them with high fidelity. Naturally, large deviations in reconstruction error would likely indicate the presence of cancer. The architecture is described below and depicted in Figure 3.

The **ConvVAE encoder** extracts hierarchical image features using four convolutional layers (`Conv2d`) with **Batch Normalization** (`BatchNorm2d`) and **ReLU activations**. Each convolutional layer reduces the spatial dimensions via strided convolutions, progressively condensing the input into a lower-dimensional feature map. The final output is flattened and passed through two fully connected layers that generate the **mean** (μ) and **log variance** ($\log(\sigma^2)$) parameters of the latent distribution.

The latent space representation is then sampled from a Gaussian distribution using the **reparameterization trick**

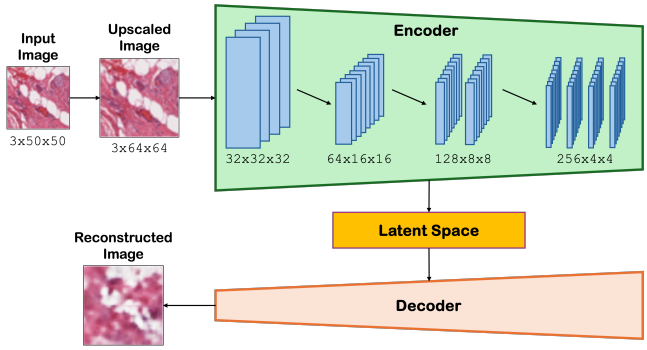


Figure 3: Vanilla Convolutional VAE Architecture

to allow gradient-based optimization. Specifically, given the mean (μ) and variance (σ^2), a sample z is computed as:

$$z = \mu + \epsilon \cdot \sigma, \quad \text{where } \epsilon \sim \mathcal{N}(0, I) \tag{1}$$

This ensures differentiability while introducing stochasticity in the latent encoding.

Finally, the **ConvVAE decoder** reconstructs the input images from the latent space using four **transposed convolutional layers** (`ConvTranspose2d`) with **Batch Normalization** and **ReLU activations**. The final output layer uses a **Tanh activation** to map the generated image with pixel values in the range $[-1, 1]$, maintaining consistency with the normalized input images.

The inspiration of the Convolutional VAE over a standard VAE with fully connected layers was the need for **spatial feature learning** in medical images. All pathological images have features embedded hierarchically, with the local cell-to-cell features at the local level and general tissue features at a higher region level [15]. Convolutional layers could be well-suited for capturing **local structures** and **hierarchical patterns**, which are essential for identifying cancer in pathology images.

D. VAE with a U-Net Encoder

To enhance the feature extraction capability of our Variational Autoencoder (VAE), we incorporate a **U-Net Encoder** based on the *ResNet-34* backbone. This encoder is designed to extract multi-scale hierarchical features from histopathological images, facilitating a more expressive latent representation. U-Net architectures have been widely adopted in medical imaging tasks, since the original U-Net, proposed by Ronneberger et. al in 2015, was used in cell segmentation [4]. As detailed in the paper, the U-Net is particularly capable of learning spatial features and capturing hierarchical patterns from pathology images. By integrating a pretrained U-Net encoder, we hypothesize that the encoder could learn a good representation of the breast tissue images.

The U-Net encoder consists of three main components: **a pretrained ResNet-34 feature extractor, a fully connected latent space transformation, and an optional weight-freezing mechanism.** The encoder utilizes a **U-Net** model implemented via the `segmentation-models-pytorch (smp)` library. The U-Net backbone is based on **ResNet-34**,

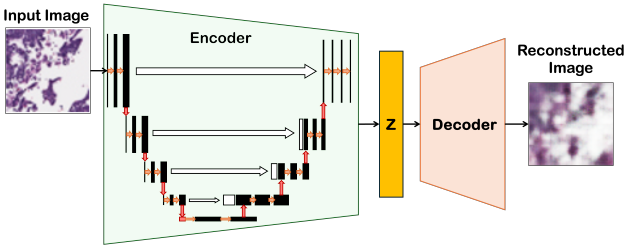


Figure 4: CVAE with Pre-trained U-NET Encoder Architecture

a deep residual network pretrained on ImageNet. The encoder outputs a final deep feature map, which serves as the input for the latent space transformation. During training, the model can either finetune the pre-trained weights or stick to frozen weights.

E. Attention-Enhanced Convolutional VAE

Building upon the previously described **ConvVAE**, we introduce an **attention mechanism** in **Attn-ConvVAE** model to enhance feature extraction and spatial relevance during encoding and decoding. Attention layers are incorporated after each convolutional operation to refine feature representations, allowing the model to dynamically focus on informative regions of histopathological images.

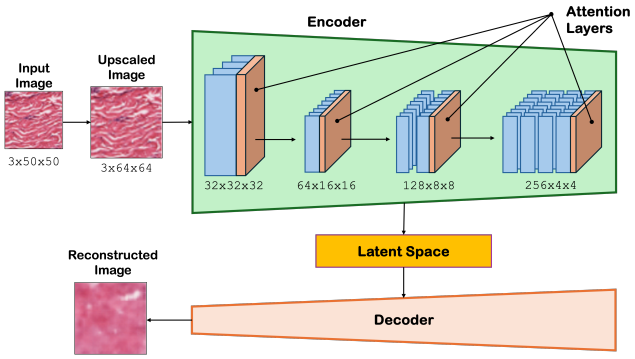


Figure 5: CVAE with Self-Attention Layers

The **attention mechanism** is integrated into the encoder and decoder of the ConvVAE. Specifically, after each convolutional layer, we apply a **self-attention module**, which enables the network to capture **long-range dependencies** and **spatial context**. This mechanism is aligned with the **non-local attention** approach, where attention weights are computed based on **query-key-value** operations.

While the base architecture remains the same as the previously discussed **vanilla ConvVAE**, the introduction of attention layers significantly alters feature learning. The ConvVAE relies solely on convolutional operations for hierarchical feature extraction, whereas **Attn-ConvVAE** augments these operations with spatial attention, allowing for more **spatially context-aware feature encoding**.

IV. Experiments

A. Dataset

The Breast Histopathology Images dataset from Kaggle was used to train and test all models. It contains 277,524 patches of size 50×50 extracted from 162 whole-mount slide images (WSI) of breast cancer specimens. The dataset is labeled with binary labels indicating the presence (`class:1`) or absence (`class:0`) of invasive ductal carcinoma (IDC). Since our application is anomaly detection, we trained our models only using the non-cancerous images, so that an increase in the loss calculation would point towards cancer.

B. Experimental Setup

The data is split into 60% training, 20% validation, and 20% test. All images (50×50) are upscaled to 64×64 using `transforms.InterpolationMode.BICUBIC`, standardizing the input shape for all models.

The model is trained using the Adam optimizer with a learning rate of 1×10^{-5} and a weight decay algorithm based on [14]. Furthermore, the models use the Wasserstein Earth Mover's Distance to determine anomalous data. To identify the optimal threshold, we employ Youden's J statistic, which maximizes the difference between true positive rate (TPR) and false positive rate (FPR), ensuring a balance between sensitivity and specificity. TPR and FPR are retrieved using the ROC curve on the labels and Wasserstein distances of the validation set. The corresponding formulations are as follows:

$$TPR = \frac{TP}{TP + FN} \tag{2}$$

$$FPR = \frac{FP}{FP + TN} \tag{3}$$

$$J = TPR - FPR \tag{4}$$

C. Evaluation Metrics

To fully assess the model's performance, we evaluate it using the following metrics:

*1) **Reconstruction Loss**:* Reconstruction loss quantifies how well the Variational Autoencoders reconstruct input images. In this case, it is computed using the Mean Squared Error (MSE):

$$\mathcal{L}_{\text{recon}} = \frac{1}{N} \sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2 \tag{5}$$

where x_i is the original input, $\hat{x}_i$ is the reconstructed output, and N is the total number of pixels.

A lower reconstruction loss indicates that the model has effectively learned a better image representation, retaining sufficient information to reconstruct images accurately. However, for anomaly detection, higher reconstruction loss is expected for out-of-distribution (cancerous) samples since the model has primarily learned to reconstruct non-cancerous samples.

*2) **Kullback-Leibler (KL) Divergence**:* KL divergence measures the difference between the learned latent distribution $q(z|x)$ and the prior distribution $p(z)$ (typically a standard Gaussian). It is computed as:

$$D_{\text{KL}}(q(z|x)\|p(z)) = \sum_i q(z_i|x) \log \frac{q(z_i|x)}{p(z_i)} \tag{6}$$

where $q(z|x)$ is the approximate posterior and $p(z)$ is the assumed prior distribution.

In VAEs, minimizing KL divergence ensures that the learned latent space follows a structured distribution, which ensures smooth interpolation and meaningful sampling. This regularization prevents overfitting as the latent variables do not become deterministic.

*3) **Overall Loss Function**:* The overall loss function of the models consists of two competing terms: reconstruction loss and KL divergence.

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \beta D_{\text{KL}}(q(z|x)\|p(z)) \tag{7}$$

The reconstruction loss ensures that the latent space captures sufficient details about the input, while KL divergence regularizes the latent space to prevent overfitting. The contribution of D_{KL} is controlled by β, which varies during KL annealing, but otherwise is 1.0.

*4) **Confusion Matrix**:* The confusion matrix provides a breakdown of the model's performance by counting the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN):

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

It helps analyze how well the model differentiates between cancerous and non-cancerous images and highlights any model tendencies to either class.

*5) **F1 Score**:* The F1 score balances precision and recall, making it especially useful for imbalanced datasets.

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

where:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

A high F1 score indicates a good balance between false positives and false negatives, which is important in medical applications. In this case, a false negative is much worse than a false positive.

*6) **Area Under the Curve (AUC)**:* AUC quantifies the ability of the model to distinguish between normal and anomalous samples across different classification thresholds. It is computed as the area under the Receiver Operating Characteristic (ROC) curve, where:

- `x`-axis represents False Positive Rate: $FPR = \frac{FP}{FP+TN}$
- `y`-axis represents True Positive Rate: $TPR = \frac{TP}{TP+FN}$

$$\text{AUC} = \int_0^1 \text{TPR}(t) d\text{FPR}(t) \tag{10}$$

AUC values close to 1.0 indicate strong discriminatory power, while values close to 0.5 suggest that the model performs no better than random guessing.

*7) **Accuracy**:* Accuracy measures the proportion of correctly classified samples over the total:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

While accuracy is generally a useful metric, it should be taken with a grain of salt, since our dataset is imbalanced with $\sim 71\%$ being non-cancerous and $\sim 29\%$ being cancerous.

V. Results

In this section, we present 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	0.41	0.64	0.63	0.65	0.64
AUC	0.53	0.70	0.60	0.50	0.53

A. *Reconstruction Loss*

The Attn-ConvVAE achieves the lowest reconstruction loss (374.57), followed by the ConvVAE (424.14). In contrast, our baseline VAE has the highest reconstruction loss (1075.94), indicating that it struggles to accurately reconstruct the input images. The Frozen CVAE-U-Net (639.11) and Unfrozen CVAE-U-Net (540.59) show moderate performance, suggesting that while the U-Net encoder improves reconstruction compared to VAE, it does not match the performance of the Attn-ConvVAE and ConvVAE models.

B. *Kullback-Leibler (KL) Divergence*

The Frozen CVAE-U-Net (20.01) and Unfrozen CVAE-U-Net (12.35) achieve the lowest KL divergence which may suggest that the U-Net encoder effectively regularizes the latent space. On the other hand, ConvVAE (261.09) and Attn-ConvVAE (392.12) have much higher KL divergence values, indicating that these models prioritize feature extraction and reconstruction over latent space regularization. The baseline VAE (48.71) falls in between all these models, showing moderate regularization.

C. *Accuracy*

The ConvVAE achieves the highest accuracy (65.58%), followed by the Attn-ConvVAE (57.89%). The baseline VAE (54.43%) performs slightly better than the Unfrozen CVAE-U-Net (53.99%) and Frozen CVAE-U-Net (51.78%), which show the lowest accuracy. This suggests that while the U-Net encoder improves latent space regularization, it does not significantly enhance classification accuracy compared to the ConvVAE and Attn-ConvVAE models. Furthermore, we see that freezing the pre-trained weights or unfreezing them provides little to no improvements. We also see that adding attention to the ConvVAE reduces the performance, which indicates that the attention module might not be beneficial.

D. *F1 Score*

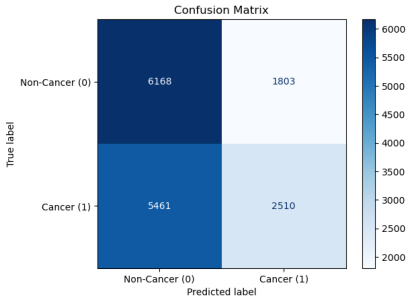
The Frozen CVAE-U-Net achieves the highest F1 score (0.65), followed closely by ConvVAE and Unfrozen CVAE-U-Net, both at 0.64, while Attn-ConvVAE trails slightly behind at 0.63. The baseline VAE falls a bit short at 0.41. This may indicate that the Frozen CVAE-U-Net, ConvVAE, Unfrozen CVAE-U-NET, and Attn-ConvVAE are relatively good at classifying cancerous and non-cancerous data while the VAE performs the weakest.

E. *Area Under the Curve (AUC)*

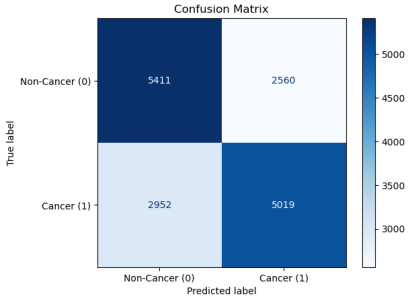
The ConvVAE achieves the highest AUC (0.70), demonstrating strong discriminatory power. The Attn-ConvVAE (0.60) and baseline VAE (0.53) follow, with the Unfrozen CVAE-U-Net (0.53) and Frozen CVAE-U-Net (0.50) showing the lowest AUC values. This further confirms that the ConvVAE and Attn-ConvVAE models are more effective at anomaly detection compared to their U-Net-based model counterparts.

F. *Confusion Matrices*

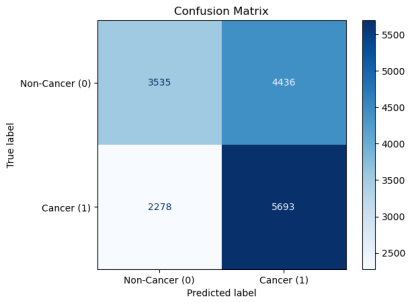
The baseline VAE (Figure a) shows a large amount of False Negatives (FN) which means it struggles to distinguish cancerous images. This is consistent with its slightly poor F1 score and AUC. Examining the ConvVAE model (Figure b), the dark color intensities along the diagonal suggest strong class separation, indicating the model's effectiveness in distinguishing between the two classes. This vast improvement is reflected with the higher F1 score and AUC. For the Attn-ConvVAE (Figure c), there's a moderate amount of predictions in the False Positive (FP) region, suggesting that the model may have a harder time distinguishing some non-cancerous images. This is also evident in the F1 score and AUC as it is lower than that of the ConvVAE model. Lastly, for the frozen CVAE-U-Net (Figure d) and unfrozen CVAE-U-NET (Figure e) models, we see that both are more likely to predict that an image is cancerous, with the unfrozen-based model performing just slightly better. This would explain the high F1 scores but low AUC of the models, as they have a high amount of True Positives (TP) but aren't able to really distinguish between cancerous and non-cancerous. Considering all the matrices, the ConvVAE appears to be the most effective for deployment.



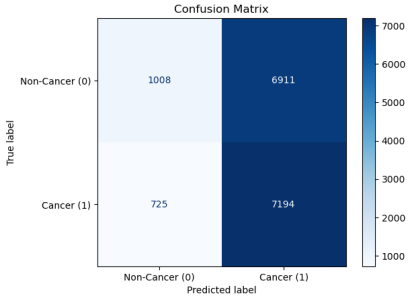
(a) VAE



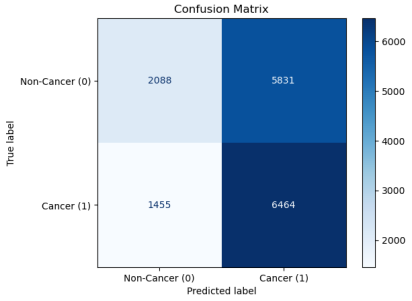
(b) ConvVAE



(c) Attn-ConvVAE



(d) Frozen CVAE-U-Net



(e) Unfrozen CVAE-U-Net

Figure 6: Confusion Matrices of Models (VAE, ConvVAE, Attn-ConvVAE, Frozen and Unfrozen CVAE-U-Net)

However, Attn-ConvVAE is also a strong contender, as it predicts fewer false negatives than ConvVAE, which is crucial for cancer detection.

VI. Contributions

A. *Diptanshu Sikdar*

Diptanshu designed and implemented the model architectures, namely the VAE, ConvVAE, U-Net-VAE, and Attention-enhanced CVAE. He trained, tuned, and conducted preliminary testing for the VAE, Vanilla ConvVAE, and Attn-ConvVAE models. Developing the models and responding their behavior was the most interesting and insighful as he learned about VAEs and their nuances on image data. In the report, he authored the Methodology section, created relevant figures, and contributed to the abstract and introduction. Besides, he worked with Jordan on the poster and set up the GitHub repository and project demo.

B. *Travis Tran*

Travis implemented evaluation metrics to assess model performance, including reconstruction loss (MSE), KL Divergence, accuracy, AUC, F1 score, and confusion matrices. He also developed code using Wasserstein distances, ROC curves, and Youden's J statistic to classify images and determine optimal thresholds. Exploring anomaly detection methods with the team was both challenging and rewarding, requiring careful consideration of suitable approaches. Researching and discussing evaluation metrics introduced him to new methodologies. In the report, Travis contributed to the experimental setup for image classification, as well as the results and conclusion sections. The results section provided quantitative comparisons and visualizations to analyze model performance, while the conclusion summarized the objectives, findings, and potential improvements.

C. *James Xu*

James collaborated with Diptanshu to develop the UNet-based Variational Autoencoder, leveraging HPC3 cluster hours to train models on over 250,000 histopathological images. Navigating Slurm job scheduling was initially challenging, but he optimized the process for efficient training. He also authored the Related Works section, distilling past and modern Anomaly Detection (AD) techniques. Discovering that the same statistical and machine-learning concepts he learned at UCI, such as Gaussian Distributions and K-Nearest Neighbors, were integral to previous AD methods put his previous knowledge into a new perspective.

D. *Jordan Yee*

Jordan collaborated with Travis to implement evaluation metrics, helping shape the project framework. His research on anomaly detection translated complex techniques into clear presentations, deepening his understanding of collaborative academic research. Jordan also contributed to key report sections, including the abstract and introduction. He co-designed the project poster with Diptanshu, condensing report content to fit the space, and organized the bibliography to credit sources.

VII. Conclusion

In this study, we explored the use of Convolutional Variational Autoencoders (ConvVAEs) for anomaly detection in breast histopathology images, comparing their performance against a baseline Variational Autoencoder (VAE) and more advanced architectures such as Attention-Enhanced ConvVAEs and U-Net-based models. Our goal was to evaluate the effectiveness of these models in distinguishing between cancerous and non-cancerous images with a focus on reconstruction-based anomaly detection.

Our results demonstrate that ConvVAEs significantly outperform the baseline VAE in terms of reconstruction loss, accuracy, F1 score, and AUC, confirming the importance of convolutional layers in capturing spatial features and hierarchical patterns in histopathology images. The ConvVAE achieved the highest accuracy (65.58%), AUC (0.70), and second highest F1 score (0.64), making it the most effective model for distinguishing between cancerous and non-cancerous images. The Attention-Enhanced ConvVAE (Attn-ConvVAE) also performed well as it achieved the lowest reconstruction loss (374.57), but its slightly lower accuracy (57.89%), F1 score (0.63), and AUC (0.60) suggest that the attention mechanism, while beneficial for reconstruction, may not significantly improve performance in this context.

In contrast, the U-Net-based models (Frozen CVAE-U-Net and Unfrozen CVAE-U-Net) showed moderate performance in reconstruction loss and achieved the lowest KL divergence, which can indicate strong latent space regularization. However, these models struggled with classification as evidenced by their lower accuracy and AUC values. Additionally, they had the tendency to overly predict images as cancerous. This suggests that while the U-Net encoder is effective for feature extraction and latent space regularization, it may not be sufficient for optimal anomaly detection in histopathology images.

The confusion matrices further support these findings, with the ConvVAE demonstrating the best balance between all cases, making it the most reliable model for deployment. The Attn-ConvVAE, while slightly less effective in overall classification, showed fewer False Negatives (FN), which is crucial for cancer detection, as missing cancerous samples can have severe consequences. This would make Attn-ConvVAE a strong candidate for deployment as well.

In conclusion, our study highlights the importance of selecting the right model architecture based on the specific characteristics of the dataset and task. ConvVAEs and Attn-ConvVAEs are particularly well-suited for histopathology image analysis, where spatial features and hierarchical patterns play a crucial role in distinguishing between cancerous and non-cancerous tissue.

Future research should explore datasets containing higher-resolution images, as U-Net and Attention-based models may yield better performance with higher image quality. Additionally, experimenting with alternative architectures, such as replacing single-head attention layers with multi-head attention layers could provide insights into model effectiveness.

Another avenue for improvement involves testing different input transformations. For denoising, techniques like Non-Local Means (NLM), Total Variation Denoising (TVD), and Wavelet-Based Denoising could be applied to enhance the clarity of the image. For contrast enhancement, methods such as Contrast-Limited Adaptive Histogram Equalization (CLAHE) and Global Histogram Equalization may improve feature extraction and model accuracy. These approaches can collectively contribute to improving model performance and generalization on many diverse datasets.

References

[1] M. Arnold, E. Morgan, H. Rumgay, *et al.*, "Current and future burden of breast cancer: Global statistics for 2020 and 2040," *The Breast*, vol. 66, pp. 15–23, 2022. DOI: 10.1016/j.breast.2022.08.010. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9465273/.

[2] H. Sung, J. Ferlay, R. L. Siegel, *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021. DOI: 10.3322/caac.21660. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33538338/.

[3] J. G. Elmore, G. Ferlay, D. Sie, A. Mafra, D. Singh, G. Longton, *et al.*, "Diagnostic concordance among pathologists interpreting breast biopsy specimens," *JAMA*, vol. 313, no. 11, pp. 1122–1132, 2015. DOI: 10.1001/jama.2015.1405. [Online]. Available: https://jamanetwork.com/journals/jama/fullarticle/2203798/.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351, arXiv, 2015, pp. 234–241. DOI: 10.48550/arXiv.1505.04597. [Online]. Available: https://arxiv.org/abs/1505.04597.

[5] H. Hecht, H. Sarhan M., and V. Popovici, "Disentangled autoencoder for cross-stain feature extraction in pathology image analysis," *Applied Sciences*, vol. 10, no. 18, p. 6427, 2020. DOI: 10.3390/app10186427. [Online]. Available: https://www.mdpi.com/2076-3417/10/18/6427.

[6] T. Authors, *Convolutional variational autoencoder*, 2024. [Online]. Available: https://www.tensorflow.org/tutorials/generative/cvae/.

[7] H. V. Guleria, A. M. Luqmani, H. D. Kothari, *et al.*, "Enhancing the breast histopathology image analysis for cancer detection using variational autoencoder," *International Journal of Environmental Research and Public Health*, vol. 20, p. 5, 2023. DOI: 10.3390/ijerph20054244. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10002012/.

[8] M. Roy, J. Kong, S. Kashyap, *et al.*, "Convolutional autoencoder based model histocae for segmentation of viable tumor regions in liver whole-slide images," *Scientific Reports*, vol. 11, p. 139, 2021. DOI: 10.1038/s41598-020-80610-9. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7794421/.

[9] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Computer Networks*, vol. 51, no. 12, pp. 3448–3470, 2007, ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2007.02.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S138912860700062X.

[10] H. Zenati, M. Romain, C.-S. Foo, B. Lecouat, and V. Chandrasekhar, "Adversarially learned anomaly detection," in *2018 IEEE International Conference on Data Mining (ICDM)*, 2018, pp. 727–736. DOI: 10.1109/ICDM.2018.00088.

[11] S. Park, K. H. Lee, B. Ko, and N. Kim, "Unsupervised anomaly detection with generative adversarial networks in mammography," *Sci Rep*, 2023. DOI: https://doi.org/10.1038/s41598-023-29521-z.

[12] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special lecture on IE*, vol. 2, no. 1, pp. 1–18, 2015.

[13] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8107–8116. DOI: 10.1109/CVPR42600.2020.00813.

[14] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *ICLR*, p. 19, 2019. DOI: 10.48550/arXiv.1711.05101. [Online]. Available: https://doi.org/10.48550/arXiv.1711.05101.

[15] R. J. Chen, C. Chen, Y. Li, *et al.*, "Scaling vision transformers to gigapixel images via hierarchical self-supervised learning," *CVPR*, pp. 16 144–16 155, 2022. DOI: 10.48550/arXiv.2206.02647. [Online]. Available: https://doi.org/10.48550/arXiv.2206.02647.