Enhancing Domain Generalization in Histopathology Image Classification through Deep Learning and Generative-AI Methods

Parastoo Sotoudeh Sharifi

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Electrical and Computer Engineering) at

Concordia University

Montréal, Québec, Canada

December 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Parastoo Sotoudeh Sharifi**

Entitled: **Enhancing Domain Generalization in Histopathology Image Classification through Deep Learning and Generative-AI Methods**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Electrical and Computer Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Wei-Ping Zhu

_____ Examiner
Dr. Yiming Xia (CSSE)

_____ Examiner
Dr. Wei-Ping Zhu

_____ Thesis Supervisor(s)
Dr. M. Omair Ahmad

_____ Thesis Supervisor(s)
Dr. M.N.S. Swamy

Approved by _____
Dr. Jun Cai, Graduate Program Director, Department of Electrical and Computer Engineering

December 16, 2025 _____
Dr. Mourad Debbabi, Dean of Faculty of Engineering and Computer Science

Abstract

Enhancing Domain Generalization in Histopathology Image Classification through Deep Learning and Generative-AI Methods

Parastoo Sotoudeh Sharifi

In recent years, advances in artificial intelligence (AI), particularly deep learning (DL), have transformed medical image analysis. Despite this progress, deploying trained deep-learning models in real-world clinical environments remains challenging due to domain shift problem—the discrepancy between the data distribution used for training and the data encountered during deployment. This issue is especially pronounced in histopathology image classification, where variations in acquisition pipelines across hospitals lead to variation in data distribution of the images. As a result, models trained on a set of source domains struggle to generalize to unseen distributions, resulting in performance degradation. This thesis addresses the challenge of domain shift in histopathology image classification by proposing two schemes that aim to improve domain generalization, each approaching the problem from a different perspective. The first scheme, referred to as PathoWAve, is more concerned with how the classifier model is trained to become robust to unseen domains. It introduces a training strategy where a single classifier model is trained along several parallel trajectories using different augmentations, and the model weights from all trajectories are averaged after each step to stabilize learning procedure. Motivated by the fact that domain shifts are bounded and continuous in the feature space, the second scheme, referred to as PathoGen, is concerned with enriching the training set with suitable synthetic images. Using a conditional stable diffusion model, PathoGen generates synthetic intermediate-domain images that lie between the original training domains. The resulting expanded training set, comprising both the original and intermediate domains, is then used to train the classifier. By combining these synthetic images with the original data, the training set becomes more continuous across domains in the feature space and accordingly increase

the likelihood of the classifier to have seen the images in the target domain. Our results indicate that both PathoWAve and PathoGen lead to significant improvements in generalization capability of the classifier models.

Acknowledgments

First and foremost, I would like to express my deep gratitude to my supervisors, Dr. M. Omair Ahmad and Dr. M.N.S. Swamy, for their invaluable guidance, unwavering support, and insightful mentorship throughout my studies. Their innovative ideas, effective leadership, and thorough reviews have significantly contributed to my academic and research growth. Their encouragement and constructive feedback have been instrumental in shaping my research direction and enhancing the quality of my work. I am profoundly grateful for the opportunity to learn and thrive under their esteemed supervision.

I extend my heartfelt thanks to the Natural Sciences and Engineering Research Council (NSERC) for their generous financial support, which has been pivotal in enabling my research. The funding provided by NSERC has allowed me to delve deeply into my studies and explore new frontiers in the field of deep learning and medical image processing.

Special thanks are due to my dear family and friends, whose unwavering support and encouragement have been a constant source of strength and motivation. I am deeply grateful for their understanding, patience, and unconditional love throughout this challenging journey. Their belief in my abilities and their constant presence have been crucial in helping me overcome obstacles and stay focused on my goals.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Network
CLIP	Contrastive Language–Image Pretraining
CPLIP	Class-Prompted Language–Image Pretraining
DA	Domain Adaptation
DG	Domain Generalization
DDPM	Denoising Diffusion Probabilistic Model
ERM	Empirical Risk Minimization
FP16	16-bit Floating Point Precision
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
H&E	Hematoxylin and Eosin
HED	Hematoxylin–Eosin–DAB
LDM	Latent Diffusion Model
MLP	Multi-Layer Perceptron
ResNet	Residual Network
TTA	Test-Time Adaptation
ViT	Vision Transformer
VAE	Variational Autoencoder
WSI	Whole Slide Image

List of Symbols

D_i	Source training domain i
$D_{i,j}$	Intermediate (synthetic) domain between D_i and D_j
H_i	Hospital or data source i
$x,\ x_0,\ x_t$	Image (pixel space); clean image; noisy image at step t
$z,\ z_t$	Latent representation (VAE latent); noisy latent at step t
c	Class label (e.g., tumor / non-tumor)
d	Domain label (e.g., hospital identifier)
B	Mini-batch of training samples
K	Number of selected synthetic samples per domain/class
M	Number of source domains
A	Number of parallel training trajectories
β_t	Variance schedule in forward diffusion
$\alpha_t = 1 - \beta_t$	Complementary variance coefficient
$\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$	Cumulative variance product
$\epsilon \sim \mathcal{N}(0, I)$	Gaussian noise sample
$\hat{\epsilon}_\theta(\cdot)$	Network's prediction of the noise
σ_t	Standard deviation of injected noise in reverse process
$t,\ T$	Current and total diffusion steps
$\mathcal{L}$	Training loss (e.g., mean squared error)
$\tau(p)$	Conditioning embedding of prompt $p = (d, c)$
C	Conditioning vector (possibly interpolated)

$\lambda \in [0, 1]$	Interpolation coefficient controlling proximity between domains
ϕ_t	Averaged global weights at iteration t
θ_{t+1}^n	Model parameters of trajectory n after iteration $t+1$
η	Learning rate
$\nabla_\phi L(\phi_t, \mathrm{AUG}_n(B))$	Gradient of loss on batch B under augmentation path n
$E(\cdot)$	Feature extractor (e.g., encoder model)
m_i	Centroidal feature of domain D_i
$m_{i,j}$	Midpoint feature between domains D_i and D_j
$d(\cdot, \cdot)$	Euclidean distance metric
$E[\cdot]$	Expectation operator
$\mathcal{N}(0, I)$	Standard normal distribution
$\|\cdot\|_2$	ℓ_2-norm

Chapter 1

Introduction

Artificial intelligence (AI) and deep learning have significantly advanced in recent years, becoming pivotal technologies in numerous fields, from natural language processing and autonomous driving to healthcare diagnostics, computer vision and image processing, financial forecasting, robotics, and creative domains like art and music generation [1].

The remarkable progress in AI can be attributed to breakthroughs in deep learning, which involves training neural networks with multiple layers to extract hierarchical patterns from data. Deep learning models have demonstrated the ability to learn from large datasets and solve highly complex tasks with impressive accuracy, making them indispensable in a variety of applications. One of the key areas where deep learning has made substantial contributions is the field of medical imaging.

In recent years, deep learning has revolutionized the field of medical image processing [2]. Leveraging complex neural network architectures, deep learning algorithms have demonstrated remarkable capabilities in extracting patterns and features from medical images, which are often imperceptible to the human eye. These models have enabled significant advancements in diagnostic accuracy, predictive analytics, and personalized medicine. Among the various sub-fields of medical imaging, histopathology, in particular, has seen profound improvements, with advancement in deep learning methods.

Histopathology is the diagnosis and study of diseases of the tissues, and involves examining tissues and/or cells under a microscope, particularly used in oncology. Traditionally, histopathological diagnosis involved manual examination of tissue slides under a microscope. However, with

technological advancements, the field has transitioned to digital pathology, where tissue sections are scanned to produce high-resolution digital images known as whole slide images (WSIs). WSIs are high-resolution, gigapixel-scale digital scans of tissue sections that provide a comprehensive view of the entire sample. WSIs enable pathologists to examine tissue morphology and cellular structures in unprecedented detail, offering insights into disease progression and aiding in accurate diagnosis. Unlike other modality images, which are typically of moderate size (e.g., 250x250 pixels), WSIs contain vast amounts of information, often requiring specialized computational techniques for processing and analysis. While WSIs have transformed histopathology, they come with unique challenges. WSIs often contain domain-specific artifacts, such as acqusition pipeline, which can hinder automated analysis. The nature of domain shifts in histopathology—such as differences in staining protocols across laboratories or changes in tissue preparation—further complicates the development of robust and generalizable models. Our research aims to tackle this unique challenges posed by WSIs, advancing diagnostic accuracy and enabling more reliable AI-driven solutions in digital pathology.

1.1 Domain Shift in Histopathology Image Classification

The intricate architecture of deep learning models involves mapping input images to increasingly abstract representations across various layers. While these representations are adept at tasks like classification, they can also inadvertently become sensitive to domain-specific details that are not essential for the task, such as minute variations in color or texture. This sensitivity is exemplified in studies around adversarial examples, where minor, often imperceptible alterations in an image can drastically mislead a model's predictions. In machine learning, and particularly in the application of deep learning to medical image analysis, the concept of domain shift represents a significant challenge. Domain shift occurs when a model trained on a specific type or domain of data (the source domain) encounters data from a different domain (the target domain) during testing or real-world application.

In the context of medical image processing, more specifically, histopathology image analysis for cancer detection, domain shift becomes particularly critical due to the high stakes involved

in accurate diagnosis [3]. Early detection of cancer plays a pivotal role in determining patient outcomes, as timely diagnosis can significantly improve the chances of successful treatment and survival. Histopathology images, which are digital representations of tissue samples, provide crucial insights into cellular structures and disease progression.

However, the analysis of histopathology images using deep learning models for disease diagnosis presents several significant challenges due to various factors that introduce variations in the images and shifts in the distribution of data used for training and testing machine learning models. These factors include: 1: Variations in Staining Techniques: Different labs may use distinct chemical stains, leading to variations in color and texture in the tissue images. These discrepancies can significantly alter the appearance of tissue structures crucial for identifying cancerous cells.

2: Differences in Imaging Equipment: Variations in microscope types, camera quality, and image resolution across different medical centers can result in distinct image characteristics, affecting the model's ability to accurately interpret new images.

3: Biological Diversity: Patient-to-patient variability and differences in disease manifestation add another layer of complexity, as the model must be able to recognize cancerous cells across a diverse range of biological presentations.

These variations result in a phenomena called "domain shift" which is a discrepancy between the training data (source domain) and the real-world data encountered during model deployment (target domain). This domain shift can lead to a marked decrease in the performance of deep learning models when they are applied to new datasets or clinical settings. These domain shifts can lead to inaccurate predictions, decreasing the ability of AI models to generalize well across diverse patient populations and healthcare settings. Inaccurate diagnosis or delayed detection could result in inadequate treatment plans, directly affecting patient care and potentially leading to life-threatening consequences for the patient.

The ability to effectively address domain shift is crucial for the practical deployment of deep learning models in histopathology image analysis. Ensuring that these models generalize well across different datasets and imaging conditions is not just a technical necessity but also a clinical imperative. Accurate and reliable models can significantly impact patient diagnosis and treatment, underscoring the importance of this research.

1.2 Approaches to Mitigating Domain Shift

Addressing domain shift involves developing strategies that enhance a model's ability to generalize across different domains without losing accuracy. This is particularly crucial in histopathology image analysis, where accurate diagnosis can be lifesaving. Techniques for mitigating domain shift fall into three main categories:

- **Domain adaptation**: This approach involves transforming data or models to bridge the gap between source and target domains. Methods range from aligning representations within a model to adversarial training that minimizes domain discrepancies [4].
- **Domain generalization**: Unlike domain adaptation, domain generalization seeks to build models that inherently possess the ability to perform well across multiple domains without requiring target domain data during training [5].
- **Test time adaptation**: This approach dynamically adapts the model to new domains during the inference phase, without retraining or prior knowledge of the target domain [6]. Test time adaptation methods adjust the model parameters, or its predictions based on the test data as it is encountered. This is different from domain adaptation and generalization as it specifically focuses on making adjustments at test time,

For this thesis we focus on second category, which is Domain Generalization for multiple domains. The methods of domain generalization can be broadly categorized into four approaches [7]:

Data-Level Generalization: These methods focus on the manipulation and generation of input data to aid the model in learning representations that are generalized. This could involve augmenting the training data with varied examples or generating synthetic data that captures the diversity of different domains.

Feature-Level Generalization: This approach centers on extracting domain-invariant features from input images. By learning a shared feature representation across multiple domains, these methods aim to improve model generalization. The focus here is on identifying and isolating features that are consistent across domains, thereby reducing the model's dependency on domain-specific characteristics.

Model-Level Generalization: Here, the emphasis is on refining the model's structure, learning process, or optimization strategies to enhance domain generalization. This could involve designing models that are inherently more robust to variations in input data or employing advanced training techniques that encourage the model to focus on generalizable patterns.

Analysis-Level Generalization: These methods assist in understanding, explaining, and interpreting the decision-making processes of machine learning models. By analyzing how models make decisions across different domains, insights can be gained into their generalization capabilities and potential biases. Our work focuses on Data-Level Generalization. In histopathology, employing these diverse methods of domain generalization can lead to the development of models that are robust against the inherent variability of medical imaging. This includes dealing with variations across different medical centers, scanners, and over time. Techniques such as data augmentation, feature extraction focused on domain-invariant characteristics, and model optimization tailored for generalization are crucial in this regard. The goal is to create deep learning models that maintain high diagnostic accuracy and reliability, irrespective of the source of histopathology images.

1.3 Overview of Existing Domain Generalization Techniques without an Apriori Knowledge of the Target Domain

Domain generalization (DG) is a critical area of research in deep learning, particularly for medical imaging where models need to generalize across different domains such as various hospitals, imaging devices, and staining procedures. This task is especially important in histopathology, where variations in imaging conditions can significantly impact model performance. Various methods have been proposed to tackle domain shift. For instance, Stochastic Weight Averaging (SWA) [8] and its extension, SWAD [9], improve generalization by averaging model weights collected from different points during a single training path. Deep CORAL [10] minimizes covariance disparity between domains, aligning feature distributions to enhance performance across diverse environments. Invariant Risk Minimization (IRM) [11] aims to learn invariant representations that hold across different domains, ensuring robustness in unseen scenarios. Group DRO (Distributionally Robust Optimization) [12] emphasizes worst-case performance, enhancing model resilience against

challenging variations. Other approaches, like FISH [13], improve domain-invariant representation learning at the feature level. LISA [14] improves robustness to domain shifts by selectively interpolating samples with the same object labels but from different domains, or with different object labels but from the same domain. This mixed strategy encourages the model to learn features consistently across domains and classes. PLDG [15] and EPVT [16] tackle effects of domain shift by incorporating two types of learnable prompts to the vision transformer model to help the model to focus on shared features across domains instead of overfitting to the domain-specific features. The domain-specific prompt captures details unique to individual domains, while the shared prompt describes common features across all domains. The difference between the methods of [16] and [15] is that the former requires the labels of the domains of the images used for training, whereas the latter determines the label of the domains through a clustering process. Data augmentation is also crucial. Methods such as StyleGAN-based approaches generate synthetic images to introduce variability into training. augmentations to test images.

1.4 Motivation and Objective

Despite the variety of methods proposed for domain generalization, several key challenges remain unresolved, particularly in histopathology image analysis. Many current approaches, such as adversarial learning and moment matching, rely heavily on direct domain alignment, which often fails to capture the subtle and inherent variations between medical institutions, staining procedures, and imaging devices. Techniques like SWA and SWAD increase the generalization to some extent, however since the averaging of the model parameters is done along one training path, these methods lack the diversity needed for strong domain generalization. Ensemble methods improve diversity by combining predictions from multiple models trained separately, each with different architectures or random initializations. This often enhances performance, but comes at the cost of high computational demand and storage requirements. Techniques like IRM and Group DRO, while effective in some cases, tend to focus on worst-case scenarios, which may not sufficiently address the nuanced domain shifts in medical imaging. Ensemble methods, which train multiple models in parallel with different weights, also provide solutions but are computationally expensive and typically

average the output of various models. This approach can enhance robustness in general cases, but is less effective for histopathology image analysis, where precision and computational efficiency are paramount. Furthermore, relying solely on augmentation strategies increases dataset diversity, but fails to substantially improve generalization across domains. LISA's approach focuses on performing interpolations on the images directly, which oversimplify complex domain variations, especially in histopathology. Consequently, multi-dimensional shifts in staining protocols or imaging characteristics cannot be effectively modeled through this simple image-level blending and mixing the features, limiting LISA's ability to consider the unseen features in between the domains and reduce its performance to effectively generalize a model to diverse, unseen scenarios. The methods of both PLDG and EPVT are able to improve the generalization capability of the classifier to some extent, but their performance depends heavily on the ability of the prompt to represent the domain specific and domain invariant features. Additionally, both methods of PLDG and EPVT are only designed for the vision transformer model architecture and they are not applicable if the classifier model is changed. Generative methods, particularly GAN-based approaches (e.g., StainGAN and CycleGAN), often suffer from mode collapse, limiting their ability to maintain domain consistency, which is crucial in medical imaging where minor deviations can impact diagnosis. Diffusion models, although more robust, lack control over the generation process, resulting in random synthetic images that fail to effectively bridge domain gaps. These methods behave similarly to augmentation techniques, expanding the dataset size without strategic control, leading to limited improvements in generalization to unseen domains.

In the view of above limitations, the objective of this thesis is to develop deep learning schemes that enhance the domain generalization capability of the classifier model used for classification of histopathology images. With this objective, we develop two different approaches for achieving the same goal, that is, the classifier is able to handle the problem of domain shift in the target domain effectively. In the first scheme, given a set of images belonging to certain domains, we develop a training strategy so that the generalization capability of the classifier is enhanced to handle the domain shift problem. A classifier is trained in multiple parallel training paths, all starting from the same initial model weights but each applying a different augmentation to the original training images. Each path runs through multiple iterations, where each iteration trains on one batch of

images from training set and as many iterations as needed to cover all batches in the training set. After every iteration, the updated weights from all paths are averaged, and this averaged weights are used to update the initial weights of all paths for the next iteration. This iterative averaging guides the model toward model weights that are consistently supported across all training paths, reducing the influence of path-specific updates and leading to a more stable and reliable optimization trajectory. The second scheme is concerned with enriching the training set with suitable synthetic images to be used for training the classifier model in order to help the classifier generalize better on unseen domains. It constructs, for every pair of original domains, a new intermediate domain and populates it with synthetic images generated by a conditional stable diffusion model. The resulting expanded training set, comprising both the original and intermediate domains, is then used to train the classifier.

1.5 Organization of the Thesis

The rest of the thesis is organized as follows:

Chapter 2: Background Material This chapter provides the background needed for the rest of the thesis. We begin with domain shift in histopathology. We then review convolutional neural networks and transformers as classifier backbones which are used in the upcoming chapters as the classifiers. Finally, we summarize the operational mechanisms and architecture of diffusion models, with emphasis on conditional stable diffusion and the training and generation mathematical principles. Together, this chapter aims to provide a comprehensive background to support the research presented in the subsequent chapters.

Chapter 3: First Proposed Method, Implementation, and Analysis Chapter 3 introduces the first proposed method, called PathoWAve, designed to improve the model's generalization capability by introducing parallel training of the identical classifier,where each training path is exposed to a different version of the training data created through various augmentation techniques. After each training iteration, the model weights are averaged accross different training paths, allowing the final classifier to capture the most stable and general patterns. This process helps the model become less sensitive to variations between hospitals and improves its overall generalization performance.

This chapter will detail the development of the method, including its theoretical basis, algorithmic structure, and implementation process. It will then present the results obtained from applying this method, providing a comprehensive analysis of its performance. The discussion will include an evaluation of the method's effectiveness, challenges encountered during implementation, and how it compares with existing techniques.

Chapter 4: Second Proposed Method, Implementation, and Analysis This chapter will introduce a second innovative approach for addressing domain shift in histopathology. The proposed method enhances feature space continuity across domains within the training set. To achieve this, PathoGen introduces intermediate domains between each pair of source domains, synthesized using a conditional stable diffusion model. These synthetic domains act as smooth transitions in the feature space, helping the classifier generalize better to unseen domains. The method's implementation details will be presented, followed by an in-depth evaluation of its effectiveness compared to existing domain generalization techniques. The chapter will conclude by analyzing the strengths and limitations of the method, along with potential improvements for future iterations.

Chapter 5: Conclusions and Future Research Directions The final chapter will summarize the key findings of the research, highlighting the contributions made to the field of domain generalization in histopathology image classification. It will discuss the practical implications of these findings, their potential application in clinical settings, and their impact on improving diagnostic accuracy. The chapter will also outline directions for future research, including ways to further improve domain generalization techniques, explore alternative approaches to diffusion models, and address the limitations encountered in this thesis. This will offer a pathway for continued advancement in the field, potentially leading to better diagnostic tools and enhanced patient care.

Chapter 2

Background Material

In this chapter, we present the background material necessary to understand the work undertaken in this thesis. We begin with an overview of domain shift, a fundamental challenge in histopathology image analysis that arises from variations across medical centers and forms the central problem addressed in this thesis. We then review convolutional neural networks (CNNs) and transformers, which serve as the backbone architectures for the methods proposed in Chapters 3 and 4. Following this, we provide an overview of diffusion models, which form the basis of the generative approach introduced in Chapter 4. Finally, the chapter concludes with a summary that connects these topics and highlights their relevance to the contributions of this thesis.

2.1 An Overview of Domain Shift

Domain shift arises when data are collected across different acquisition pipelines, causing the image distribution to change. In the Camelyon17-WILDS benchmark, each *domain* corresponds to a hospital: models are trained on Hospitals 1–3 and evaluated on unseen hospitals (validation on Hospital 4 and test on Hospital 5). The visual differences across hospitals are illustrated in Figure 2.1, adapted from Koh et al. [17], and motivate the domain-generalization focus of this thesis.

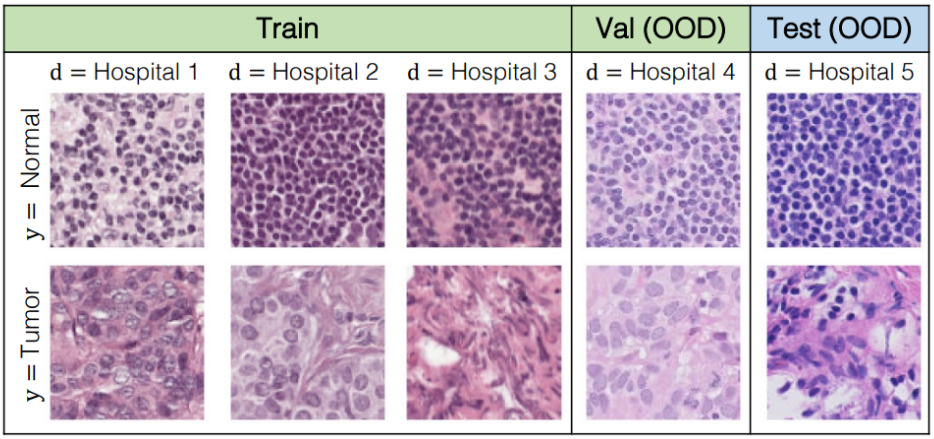

Figure 2.1: The Camelyon17-wilds dataset comprises tissue patches from different hospitals. The goal is to accurately predict the presence of tumor tissue in patches taken from hospitals that are not in the training set. In this figure, each column contains two patches, one of normal tissue and the other of tumor tissue, from the same slide.

2.2 Convolutional Neural Networks

Convolutional neural networks (CNNs) have played a transformative role in deep learning, achieving outstanding results in both computer vision and, more recently, natural language processing. The first widely recognized CNN, LeNet, was introduced by LeCun et al. [18], and was designed to recognize handwritten digits. With advances in computing power, particularly the availability of GPUs capable of parallel processing, researchers were able to increase the depth of CNNs significantly. Deeper architectures such as VGG16 [19] and ResNet [20] have become well-established benchmarks in computer vision. VGG models are representative plain networks, composed of stacked convolutional, pooling, and fully connected layers without shortcut connections. Although increasing their depth can enhance representational power, such plain networks suffer from vanishing gradients, often leading to degraded performance [20]. Residual Networks (ResNets) address this issue by introducing skip connections, enabling stable training of much deeper models [20], as illustrated in Figure 2.2. In this thesis, ResNet-50, an extension of ResNet-34 with bottleneck blocks, is used as the backbone for histopathology image classification. CNNs have been widely applied across many areas of computer vision and have become one of the most established architectures for image classification tasks [19, 20]. Their ability to automatically extract hierarchical features from raw images has made them especially effective in medical image

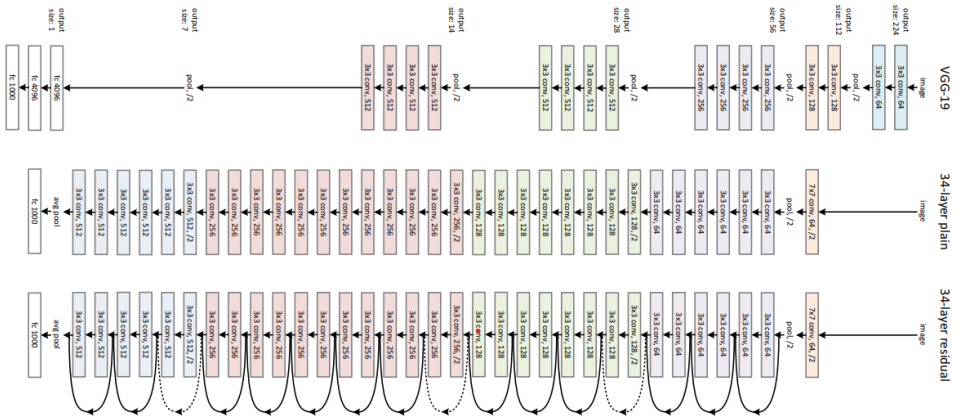

Figure 2.2: Comparison between VGG-19 (left) and a 34-layer Residual Network (ResNet-34, right). The figure illustrates how residual connections enable the training of deeper networks.

analysis, including histopathology images.

A standard CNN architecture typically contains three main building blocks: convolutional layers, pooling layers, and fully connected layers. Each component has a distinct role. Convolutional layers perform feature extraction by applying multiple learnable filters of size $D \times H \times W$, where D is the depth of the filter, and H and W are the spatial dimensions. Pooling layers reduce the spatial resolution of feature maps, lowering computational complexity while preserving important information. The most common pooling operations are max pooling and average pooling. Fully connected layers are generally used at the end of the network to perform tasks such as image classification, clustering, or speech recognition. In classification problems, the output of the final fully connected layer is typically passed through a Softmax function to obtain class probabilities.Modern CNNs often include additional operations, such as batch normalization and layer normalization, to stabilize training and accelerate convergence.

2.3 Transformers

The transformer architecture was originally introduced for natural language processing tasks, where it quickly became the dominant framework due to its ability to capture long-range dependencies [21]. The general architecture consists of two main components, an encoder and a decoder.

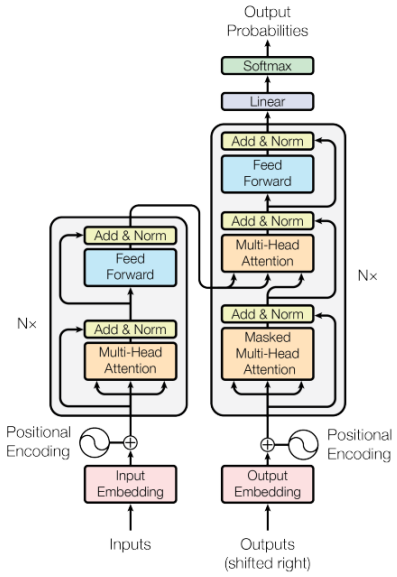

Figure 2.3: Transformer encoder–decoder architecture.

The encoder is built from six identical layers, each containing a multi-head self-attention mechanism and a position-wise feed-forward network. Residual connections and layer normalization are applied around each of these sub-layers [20, 22].

Same as the encoder, the decoder is composed of six identical layers. In addition to the two operations used in the encoder, each decoder layer introduces a third multi-head attention block that incorporates information from the encoder output. Residual connections and normalization are again employed, and the decoder's self-attention mechanism is slightly modified to improve prediction accuracy. The complete transformer architecture is illustrated in Figure 2.3, adapted from the original work [21]. Inspired by their success in NLP, transformers have been adapted to computer vision, most prominently through the Vision Transformer (ViT) [23], which divides an image into non-overlapping fixed-size patches, embeds them with positional information, and feeds the sequence as input tokens to a Transformer encoder for classification. An overview of this ViT pipeline is shown in Figure 2.4.

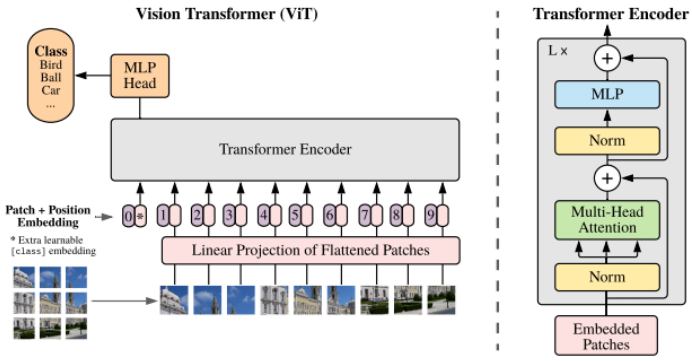

Figure 2.4: ViT Model overview. Image is splitted into fixed-size patches, linearly embed each of them, add position embeddings, and feed the resulting sequence of vectors to a standard Transformer encoder. In order to perform classification, use the standard approach of adding an extra learnable "classification token" to the sequence. The illustration of the Transformer encoder was inspired by Vaswani et al. (2017).

2.4 Diffusion Models

Diffusion models are a class of generative models capable of producing high-quality images while avoiding issues such as mode collapse and heavy sensitivity to hyperparameter tuning. They learn the training data distribution and synthesize new samples from it. Training proceeds by progressively adding noise to clean samples in a forward process and learning to invert this corruption to recover the original data in a reverse denoising process [24]. In the DDPM formulation, this is formalized with a variance schedule for the forward corruption, and the denoiser is trained to predict the injected Gaussian noise using a simple mean-squared error objective [25]. After training, the model can start from pure noise and iteratively denoise to generate images from the learned distribution. Architecturally, these models often use an encoder and a symmetric decoder built from residual blocks and attention modules.

Stable Diffusion is a diffusion variant designed to reduce computation for high-resolution synthesis by carrying out the denoising in a learned latent space rather than in pixel space. In this setup, an encoder compresses the image to a low-dimensional latent, the denoising network operates on this latent, and a decoder reconstructs the image at the end. The encoder typically includes several downsampling stages with residual convolutional layers and a central attention module, while the decoder mirrors this with corresponding upsampling stages. A conditional diffusion model steers

generation using conditioning signals (e.g., text or visual embeddings) [26] . In this thesis, to produce images whose distributions are shifted relative to the original training domains, we employ a conditional Stable Diffusion model [26]. The architecture of the stable diffusion model is shown in Figure 2.5.

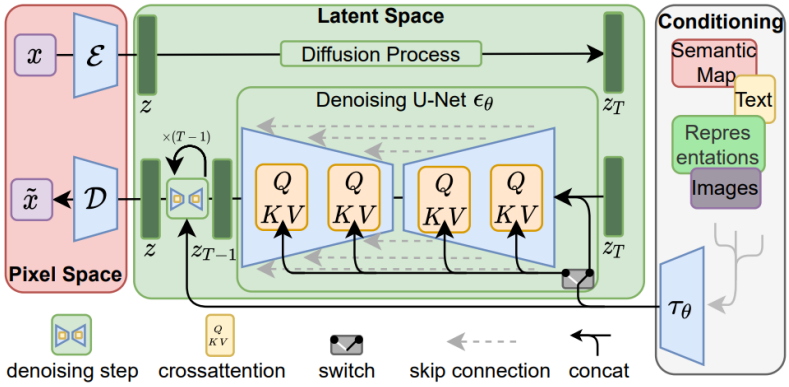

Figure 2.5: Stable Diffusion Model Architecture.

2.4.1 Training Phase of Diffusion Model

We follow the DDPM formulation of Ho et al. [25]. Let $\{\beta_t\}_{t=1}^{T}$ be a variance schedule, $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. The training phase consists of a forward and reverse process. In the forward pass, we pick a random timestep t and corrupt a clean image x_0 with Gaussian noise. A clean image x_0 is progressively corrupted.

$$x_t \;=\; \sqrt{\bar{\alpha}_t}\, x_0 \;+\; \sqrt{1-\bar{\alpha}_t}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I), \quad t \sim \text{Uniform}\{1, \ldots, T\}. \tag{1}$$

In the reverse path, the model is trained to predict the noise $\epsilon_\theta(x_t, t)$ of the sample x_t at timestep t according to the following loss function (ℓ_2 loss):

$$\mathcal{L} \;=\; E_{x_0,\epsilon,t}\Big[\, \|\epsilon \;-\; \epsilon_\theta(x_t, t)\|_2^2 \,\Big]. \tag{2}$$

2.4.2 Generation Phase of Diffusion Model

In the generation phase, we start from pure noise $x_T \sim \mathcal{N}(0, I)$,, and iteratively denoise the sample using the trained diffusion model to generate a new sample:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) + \sigma_t z, \qquad z \sim \mathcal{N}(0, I), \tag{3}$$

where a common choice is $\sigma_t = \sqrt{\beta_t}$. The final sample x_0 is the newly generated sampled by the diffusion model which has the distribution similar to that of the original training samples.
In stable diffusion [26], the same equations are applied in a latent space z of a learned autoencoder; the generated latent z_0 is decoded to the image $x_{\text{gen}} = \text{Dec}(z_0)$.

2.5 Summary

This chapter has presented the relevant background material necessary for understanding the work of this thesis. The chapter started with a discussion of domain shift in histopathology and the motivation for domain generalization, followed by brief reviews of convolutional neural networks and transformers as the classifier backbones used in Chapter 3 , 4. Finally, the chapter ends with a discussion of diffusion models which underpins the generative approach of Chapter 4.

Chapter 3

PathoWAve: A Deep Learning-based Weight Averaging Method for Improving Domain Generalization in Histopathology Images

3.1 Introduction

In this chapter, we propose a domain generalization scheme, known as PathoWAve, which introduces a new training strategy designed to enhance the generalization capacity of the classifier to handle the domain shift problem [27]. PathoWAve stabilizes the learning process by training an identical classifier model along multiple parallel training paths, where each path exposed to differently augmented versions of the training data. Each path runs through multiple iterations, where each iteration trains on one batch of images from the training set and as many iterations as needed to cover all batches in the training set. After every iteration, the updated weights from all paths are averaged, and this averaged weights are used to update the initial weights of all paths for the next

iteration. This iterative procedure enables the model to learn representations that are broadly representative of the entire training distribution while benefiting from the diversity introduced through augmentation.

In contrast to many domain generalization methods that which often rely on training several distinct model architectures to capture diverse feature patterns, PathoWAve uses a single architecture, reducing computational cost. Additionally, PathoWAve is its ability to generalize well to unseen data without requiring any adjustment during test time and without any access to information from test data during training. This characteristic makes the method suitable for real-world clinical applications

3.2 Proposed Method

PathoWAve is a multi-source domain generalization framework specifically designed to address domain shift challenges in histopathology images. The proposed approach aims to leverage multiple source domains and build a model that can generalize well to these unseen target domains. To achieve this, PathoWAve adopts a cyclical training regime, which integrates advanced weight averaging techniques and domain-specific augmentation strategies. The goal is to improve robustness to domain shift arising from differences in acquisition pipelines across medical centers. In the following subsection we will go through the details of each of the main components of the proposed scheme. An overview of the proposed method is shown in Figure 3.1.

3.2.1 Multi-Trajectory Training Approach

One of the core innovations of PathoWAve is its multi-training trajectory strategy. Rather than training a single model on the entire dataset, we train multiple identical models concurrently. These models start from a common initialization point within the loss landscape but follow different trajectories due to the unique augmentations applied to each model's data. This ensures that the models explore diverse representations of the data, improving their ability to generalize to new domains.

Mathematically, the training process for each model i can be described as:

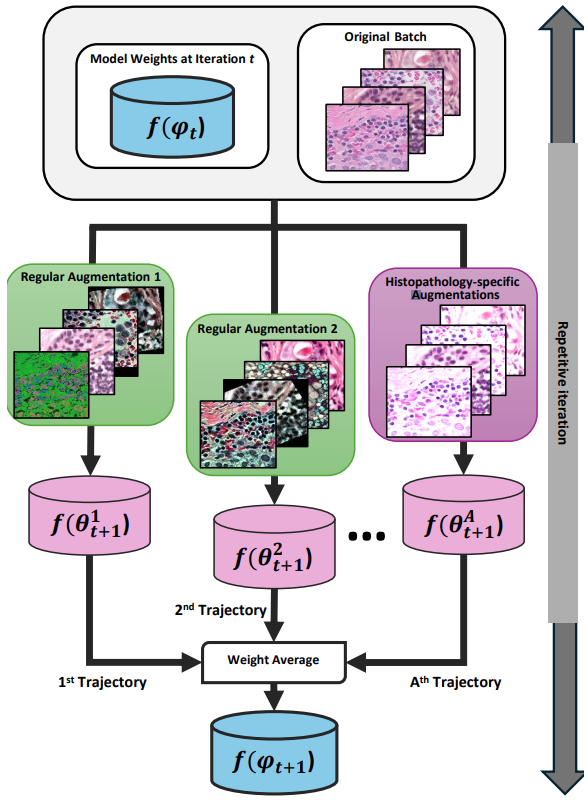

Figure 3.1: Overview of the PathoWAve training process. Multiple training trajectories run in parallel, each with distinct augmentations (regular and histopathology-specific). After each iteration, weights are averaged to promote flatter minima and improve generalization.

$$\theta^n_{t+1} = \phi_t + \eta \cdot \nabla_\phi L(\phi_t, \text{AUG}_n(B)), \tag{4}$$

where θ^n_{t+1} represents the model parameters for the n-th model after iteration $t+1$, η is the learning rate, and $\nabla_\phi L(\phi_t, \text{AUG}_n(B))$ is the gradient of the loss function L evaluated on the augmented batch $\text{AUG}_n(B)$, which corresponds to the specific augmentation strategy n applied to batch B. This concurrent training process allows each model to be exposed to a diverse range of augmented data, ensuring that the model learns a wide variety of features from different representations of the input space. At the core of the PathoWAve framework is the cyclical weight averaging mechanism during training time, which integrates the weights of each model after every training iteration. This weight averaging process is essential for guiding the models toward flatter minima in the loss landscape, which is known to improve generalization performance.

The concept of flatter minima refers to regions in the loss landscape where the loss function changes slowly around the minima, leading to more robust models. Models that converge to sharp minima (regions where the loss increases steeply) are often highly sensitive to small changes in the input data, which can lead to poor generalization in new domains. In contrast, models that find flatter minima are more stable and less sensitive to variations in the data, resulting in better performance on unseen domains.

In PathoWAve, after each training iteration, the weights of all model instances are averaged to ensure convergence toward flatter minima. The weight averaging process is defined as:

$$\phi_{t+1} = \frac{1}{A} \sum_{n=1}^{A} \theta_{t+1}^{n}, \tag{5}$$

where ϕ_{t+1} represents the averaged weight vector, θ_{t+1}^{n} are the parameters of the n-th model at iteration $t+1$, and A is the total number of models. This averaged weight vector is then used as the initialization point for the next iteration. By averaging the weights of the models, PathoWAve ensures that the models explore different parts of the loss landscape but ultimately converge to a shared, flatter solution. This approach enhances the model's robustness to domain shifts, ensuring that the model generalizes better to new, unseen domains.

3.2.2 Comprehensive Augmentation Strategies

Another key element of PathoWAve is its combination use of both general and histopathology-specific data augmentation techniques. Augmentation plays a critical role in ensuring that the model is exposed to a wide variety of data transformations, allowing it to learn invariant features that generalize across different domains.

General Augmentations: Techniques such as *AutoAugment* and *RandAugment* [28] are used to introduce common image transformations like flipping, rotation, and scaling. These augmentations ensure that the model learns features that are robust to the general variability found in medical imaging, helping it adapt to a broad range of real-world imaging conditions.

Histopathology-Specific Augmentations: To address the specific challenges in histopathology,

we use augmentations such as *HEDJitter* [29], which simulates variability in tissue staining protocols. Histopathology images are typically stained using hematoxylin, eosin, and diaminobenzidine (HED), and the intensity of these stains can vary significantly across laboratories. *HEDJitter* operates in the HED color space, where each stain channel is perturbed with random noise to mimic these variations. Mathematically, this is represented as:

$$I'_{\text{HED}} = I_{\text{HED}} + \mathcal{N}(0, \sigma_{\text{HED}}), \tag{6}$$

where I_{HED} is the image in the HED color space, and $\mathcal{N}(0, \sigma_{\text{HED}})$ represents the Gaussian noise added to each stain channel. The augmented image is then converted back to the RGB space, providing a diverse set of training samples that reflect the real-world variability in staining protocols. This ensures that the model learns features that are invariant to staining variations, which is critical for generalizing across different medical centers.

By integrating the cyclical weight averaging strategy with comprehensive data augmentation, PathoWAve creates a robust framework for domain generalization. The parallel training trajectories, each exposed to different augmentations, ensure that the model explores diverse representations of the data. The weight averaging mechanism then facilitates convergence toward flatter minima, which enhances the model's generalization capabilities.

In each training cycle, the model is trained on augmented data batches from the union of all source domains, denoted as $\sum_{i=1}^{M} D^i_{\text{source}}$. After every iteration, the weight averaging strategy is applied, creating a unified weight set that serves as the starting point for the next iteration. This structured, iterative refinement process ensures that the model is exposed to a broad spectrum of data variations while also ensuring convergence toward a solution that generalizes well across unseen domains. The combination of general and histopathology-specific augmentations enhances the model's adaptability and generalization, while the cyclical weight averaging ensures stability and robustness. This dual strategy allows PathoWAve to set a new benchmark in domain generalization for histopathology image analysis, ensuring reliable performance in real-world clinical settings.

3.3 Experimental Results

In this section, we present the comprehensive evaluation of the proposed PathoWAve framework on the Camelyon17 WILDS dataset. We first describe the dataset used for the evaluation of the proposed scheme, followed by key implementation details. We then compare the proposed PathoWAve scheme against other existing domain generalization methods. Finally, an ablation study quantifies the contribution of each component, showing that multi-trajectory training with cyclical weight averaging and histopathology-specific augmentations will significantly boost the accuracy performance of the classifier.

3.3.1 Dataset

The dataset used in our experiments is the Camelyon17 WILDS dataset, which is specifically designed for evaluating domain generalization in histopathology image analysis. This dataset consists of WSIs of lymph node sections collected from five different medical centers. The diversity in staining techniques, scanner devices, and imaging protocols across these medical centers introduces significant domain shifts, making the dataset an ideal benchmark for testing the generalization capabilities of models on unseen domains.

The Camelyon17 dataset is partitioned by the origin of the medical centers to simulate real-world domain generalization tasks. For training, the data includes WSIs from three medical centers, which provide a total of 302,436 image patches derived from 30 WSIs. These patches are used for both training and in-domain validation (id val). The in-domain validation set consists of 33,560 patches from these same centers. The validation set (val) and testing set are drawn from two different unseen medical centers. The validation set contains 34,904 patches extracted from 10 WSIs, while the test set consists of 85,054 patches from another 10 WSIs.

This partitioning strategy ensures that the models are trained on data from certain centers and evaluated on unseen centers, simulating a domain shift between training and testing data. This makes the dataset suitable for measuring how well models generalize to new domains without explicit access to test-time data during training.

3.3.2 Implementation Details

For our experiments, we utilized the ResNet50 architecture, a deep convolutional neural network known for its strong performance in image classification tasks. We trained the model on NVIDIA V100 32 GB GPUs, which provided sufficient computational power for handling the large-scale data and extensive augmentations. The learning rate was set to 2×10^{-5} and the batch size was set to 128 for all experiments. To further improve the model's robustness to staining variations commonly observed in histopathology images, we incorporated several augmentation techniques. Specifically, HEDJitter augmentation with a jitter strength of 0.05 was applied.

3.3.3 Comparison with other schemes

In this section, we present the comprehensive evaluation of the proposed PathoWAve framework. Our evaluation demonstrates the exceptional generalization capability of PathoWAve across domain shifts within histopathology images. We compare PathoWAve against a variety of state-of-the-art domain generalization (DG) methods.

Tables 3.1 and 3.2 summarize the performance comparison of PathoWAve with existing domain generalization methods on ResNet50 and Vision Transformer (ViT) classifier, respectively. The results clearly illustrate that PathoWAve achieves superior performance, outperforming both ResNet-based and Vision Transformer (ViT)-based architectures.

Table 3.1 presents the test-time classification performance of various domain generalization methods on the Camelyon17 WILDS dataset using a ResNet-based classifier. PathoWAve achieves the highest accuracy of 94.36%, surpassing the second-best method, TestTimeI2I [37] (94.0%), by a margin of 0.36%. It is worth mentioning that in contrast to TestTimeI2I [37], which increase the test time complexity, our method has no additional test time overhead. This consistent improvement demonstrates the superior generalization capability of PathoWAve under convolutional backbones. Similarly, as shown in Table 3.2, PathoWAve attains an accuracy of 94.89% with the ViT-based classifier, outperforming the second-best method, EPVT [41] (86.4%), by 8.49%. These results collectively highlight PathoWAve's robust and architecture agnostic ability to enhance domain generalization performance across both convolutional and transformer-based models.

Method	Test Accuracy (%)
CORAL [10]	59.5
IRM [11]	64.2
Group DRO [12]	68.4
DomainMix [30]	69.7
MMLD [31]	70.2
ERM [17]	70.3
VREx [32]	71.5
IB-IRM [33]	68.9
FISH [13]	74.7
LISA [34]	77.1
FuseStyle [35]	90.5
STRAP [36]	93.7
StarGANv2 [37]	76.4
TestTimeI2I [37]	94.0
PathoWAve (Proposed)	**94.36**

Table 3.1: Comparison of PathoWAve with state-of-the-art domain generalization methods using a ResNet50 backbone on the Camelyon17 WILDS dataset.

Method	Test Accuracy (%)
CORAL [10]	71.8
IRM [11]	75.0
ERM [38]	73.1
SelfReg [39]	70.4
DANN [40]	83.5
PLDG [38]	84.3
EPVT [41]	86.4
PathoWAve (Proposed)	**94.89**

Table 3.2: Comparison of PathoWAve with state-of-the-art domain generalization methods using a ViT-Base backbone on the Camelyon17 WILDS dataset.

3.3.4 Ablation Studies

We performed a detailed ablation analysis to investigate the impact of different augmentation strategies and the number of independent training trajectories on PathoWAve's performance. Table 3.3 summarizes the results of this analysis, revealing how combinations of augmentations and training trajectories affect the model's generalization capability on the Camelyon17 WILDS dataset. Initially establishing a baseline with the ERM method, which utilizes a single training trajectory without weight averaging, yielded a test accuracy of 70.03%. The introduction of PathoWAve with dual augmentation strategies significantly enhances model performance, highlighting the method's

responsiveness to diverse training signals. Notably, combinations involving two augmentations, particularly AutoAugment with HEDJitter, demonstrated remarkable improvements, achieving a test accuracy of 94.20%. This underscores the critical role of HEDJitter, a histopathology-specific augmentation, in bolstering the model's generalization capability across unseen domains. Further

Method	**# Independent Trajectories (Augmentations)**	**Test (%)**
ERM	1 (baseline with no weight averaging)	70.3
PathoWAve	2 (AutoAugment, RandomAugment)	92.53
PathoWAve	2 (RandomAugment, HEDJitter)	92.98
PathoWAve	2 (AutoAugment, HEDJitter)	94.20
PathoWAve	3 (AutoAugment, RandomAugment, AutoRandomRotation)	89.80
PathoWAve	3 (AutoAugment, RandomAugment, RandomGaussBlur)	88.91
PathoWAve	3 (AutoAugment, RandomAugment, RandomAffine)	91.53
PathoWAve	3 (AutoAugment, RandomAugment, HEDJitter)	94.36

Table 3.3: Ablation analysis of PathoWAve method on the Camelyon17 WILDS dataset.

exploration with three augmentations revealed varying degrees of success. While adding AutoRandomRotation, RandomGaussBlur, or RandomAffine to the AutoAugment and RandomAugment mix led to lower test accuracies compared to dual-augmentation setups, the incorporation of HEDJitter alongside AutoAugment and RandomAugment within a three trajectory framework achieved the highest performance at 94.36%. This pinnacle result not only signifies the optimal augmentation combination but also establishes PathoWAve as the state-of-the-art in domain generalization for histopathology images. It is worth mentioning that our proposed method's training time is A times that of traditional one-trajectory methods, as we perform A augmentations in parallel per iteration before weight averaging. In our method A is equal to 3, due to the fact that we have three training trajectories. Importantly, this overhead is only during training; the testing time remains the same as other methods since we use the averaged weights for evaluation.

3.4 Summary

This chapter has been presented PathoWAve, a novel training strategy designed to improve domain generalization in histopathology image classification.

The approach combines parallel training trajectories with different augmentations and a weight-averaging process to build a more stable and robust model. The approach integrates multiple parallel training paths, each using the same classifier model but using distinct data augmentations. Each path trains for several iterations, with each iteration processing one batch of images. After every iteration, the updated weights from all paths are averaged and used to reinitialize all paths for the next iteration. The ablation study demonstrated that the combination of multiple training paths and diverse augmentation strategies played a major role in improving the overall performance. Experiments on the Camelyon17-WILDS dataset showed that PathoWAve achieved the highest test accuracy compared to the existing domain generalization state-of-the-art methods in the literature.

Chapter 4

PathoGen: A Generative Diffusion-Based Domain Generalization Scheme for Histopathology Image Classification

4.1 Introduction

The difference between acquisition pipelines across hospitals is resulting in domain shift problem and degrade the accuracy of the histopathology image classification. Enhancing the generalization capability of the classifier is therefore a key objective. Within this objective, in Chapter 3, we developed a novel training strategy that stabilizes learning procedure of the model to enhance the generalization capability of the classifier. In this chapter, motivated by the fact that domain shifts are bounded and continuous in the feature space, we propose a scheme, referred to as PathoGen, which focus on generating synthetic images belonging to intermediate domains between the original training domains [42]. Adding the new intermediate domains to the training set will increase the continuity of the feature space and accordingly increase the likelihood of the classifier to have seen

the images in the target domain. PathoGen follows a two-stage process. In the first stage, a conditional stable diffusion model is trained on images from the original domains and used to generate synthetic samples that lie between pairs of these domains. This expands the diversity of the training data beyond what is observed in the original set.

In the second stage, feature representations are extracted for both original and synthetic images. For each pair of original domains, the midpoint between their feature-space centroids defines the corresponding intermediate domain. Synthetic images closest to this midpoint are selected to populate that intermediate domain. The resulting expanded training set, comprising both the original and intermediate domains, is then used to train the classifier. PathoGen improves generalization capability of the classifier model to unseen target distributions without requiring access to target data during training and without relying on the classifier architecture.

4.2 Proposed Method

A histopathology acquisition pipeline comprises several steps including tissue fixation and processing, tissue sectioning into thin slices, staining (such as H&E), mounting the sample on a slide, and finally digitizing it with a whole-slide scanner [43]. The pipelines of most hospitals follow these steps, but some details differ, such as the thickness of the section, the type and timing of stains, or the brand and settings of the scanner. Studies that compare slides from different hospitals show that these differences mainly affect the appearance of the images, for example, color balance, contrast, or sharpness, while the underlying tissue structures remain the same [44]. Prior work has shown that these differences manifest as bounded, continuous shifts in color/contrast statistics and embeddings rather than disjoint pathology, implying that real-world target domains lie near source domains in feature space [45]. In this section, we develop a domain generalization framework, referred to as PathoGen, to improve the robustness of a classifier in histopathology image classification across unseen domains. Our proposed scheme is aimed at creating new domains of synthetic images which, when included in the original training set, make the feature space originally comprising the source domains more continuous. The proposed scheme, shown in Figure 4.1, comprises two stages. In the first stage, we exploit the capability of a conditional stable diffusion model to generate synthetic

images which belong to the intermediate domains between each pair of the original training set. In the second stage, we develop a method for forming the intermediate domains using the generated images and expanding the original training set by including in it images from both the original domains and the newly created intermediate domains. The tasks carried out in the two stages are done with a view to enhance the generalization capability of a classifier model by training it with the expanded training set thus obtained. The two stages of the proposed scheme are now explained in detail in the following two subsections and shown in Figure 4.1

4.2.1 Generation of Candidate Synthetic Images

Since our objective in this subsection is to generate candidate synthetic histopathology images that belong to domains that are positioned between each pair of the original domains, we first describe the architecture of the model to generate candidate images, then explain the procedure for its training and finally discuss how this trained model is utilized to generate new synthetic images.

Architecture of Image Generating Model

Diffusion model [24] is a type of generative model which is capable of producing different types of high-quality images without running into issues like mode collapse or heavy dependence on finely tuned hyperparameters. Diffusion models learn the distribution of the training data set and generate new samples that belong to the same distribution. Diffusion models take a unique training approach, by progressively adding noise to a sample of the training data in the forward path and then optimising the model by reversing this process to recover the original data from the diffused image. Once the model has been trained, it becomes capable of generating from noise samples new images that belong to the same distribution as that of the samples in the training set. The architecture of this model has an encoder and a symmetric decoder each consisting of a certain number of residual blocks and attention modules. The stable diffusion model [26] is a type of diffusion model specifically designed to reduce computational cost for the generation of high-resolution images by caring out the denoising processes both in the training and generation phases in the latent space instead of the pixel domain. In the architecture of this model, the main diffusion model is preceded by an encoder module to compress the input image and is followed by a decoder

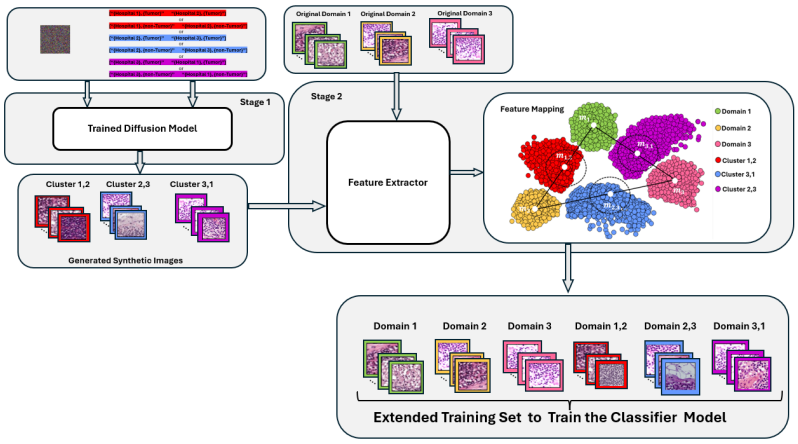

Figure 4.1: Overview of the PathoGen Framework for Domain Generalization in Histopathology Image Classification. The Framework Consists of Two Main Phases: (Left) Candidate Synthetic Image Generation using a Conditional Diffusion Model, and (Right) the Formation of Intermediate Domains and Generalization Enhancement of Classification. The Details of the Trained Diffusion Model box is presented in Figure 4.2.

to decompress the low dimensional image generated by the main model. The encoder module at the input consists of certain number of down sampling blocks each containing certain number of residual convolutional layers, with an attention module in the middle. The structure of the decoder at the output is symmetric to that of the encoder with the down sampling blocks replaced by up sampling blocks. The conditional diffusion model is a version of diffusion model in which the model is conditioned to control the distribution of the images generated. The distribution of the images to be generated is passed on to the network using a text message or directly through a visual message through an embedding mechanism. In our proposed scheme, for generating the images with distributions shifted from that of the original training images, for training of classifier model, we used a conditional stable diffusion model [26].

Training of the Image Generating Model

We train the conditional stable diffusion model on all the images of all training domains so as to learn the distribution of the training domains. Since our goal is to generate domain-aware histopathology images, we incorporate conditional information in terms of both the domain and the class of the training images during the training stage of the model. In a conditional stable diffusion

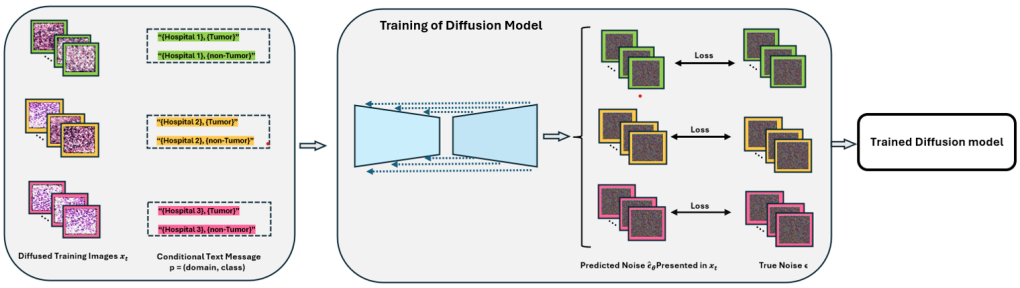

Figure 4.2: Training procedure of the conditional stable diffusion model. The model is trained on diffused training images x_t from multiple domains (Hospital 1–3) along with their associated class labels (Tumor / Non-Tumor). Conditional information (d, c), representing domain and class, is embedded into a vector $\tau(p)$ and provided to the diffusion model. During training, the model learns to predict the noise $\hat{\epsilon}_\theta$ added to latent representations z_t, minimizing the error between predicted noise and true noise ϵ. The resulting trained diffusion model captures domain- and class-aware distributions of histopathology images.

model this is achieved by inputting to it an embedding vector C that contains the information on the domain and the class of the diffused training image that is input to the diffusion model. The embedding vector C = $\tau(p)$ is produced by a language model by prompting it by the text message p = (d, c), where d and c represent, respectively, the domain and the class names of an image which also has been shown in Figure 4.2. For training the model, first, for each training image x, a latent (compressed) representation z is obtained using a VAE encoder. The latent image z then undergoes a forward diffusion process, employing a sample of Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$, consisting of t timesteps, where t is randomly sampled from the range $[0, T]$, with T being the maximum number of diffusion steps to be used for diffusing any image. At each of the timesteps from 1 to t, the noise sample ϵ is used to diffuse the clean latent z according to a cosine noise schedule to finally obtain the noisy latent representation z_t of z at the end of t-th timestep. The cosine noise schedule determines how much noise is added to the latent representation of the previous timestep to obtain the representation at the current step. In practice, z_t can be obtained directly by using the following equation [25].

$$z_t = \sqrt{\alpha_t} \cdot z + \sqrt{1 - \alpha_t} \cdot \epsilon \tag{7}$$

where $\alpha_t = \frac{f(t)}{f(0)}$, with f(t) given by

$$f(t) = \cos^2\left(\frac{\frac{t}{T} + s}{1 + s} \cdot \frac{\pi}{2}\right) \tag{8}$$

In Eqn. (2), the quantity s is a small constant used to avoid numerical instability, sharp gradients, and poor learning at early timesteps by gently shifting the cosine schedule to the right. The value of s suggested in [46] is s = 0.008.

The model is trained to predict the noise sample ϵ used to diffuse the original latent image z into z_t. The predicted noise, denoted by $\hat{\epsilon}_\theta(z_t, t, \tau(P))$, is the model's prediction of the noise sample ϵ used in the forward process. The loss function used for training the model is given by:

$$L_{\text{stable diffusion}} = E_{x,p,\epsilon,\, t}\left[\|\epsilon - \hat{\epsilon}_\theta(z_t, t, \tau(p))\|_2^2\right] \tag{9}$$

where x represents a training image, p is the associated conditioning information on the domain and the class of x, which is encoded by the function $\tau(p)$, ϵ is a sample of Gaussian noise used to obtain the diffused image latent z_t from the latent image z employing t steps during the forward diffusion process, and $\hat{\epsilon}_\theta(z_t, t, \tau(p))$ is the model's prediction of ϵ. In Eqn. 3, the L2 norm, denoted by $\|.\|_2$, is first computed on the individual images in a batch; then, the expectation operation, denoted by E, obtains the average of the squared L2 operation over all the images in the batch. The loss function as defined by Eqn. 3, measures how close is the predicted noise sample $\hat{\epsilon}_\theta$ to the true noise sample ϵ used during the forward diffusion process.

Image Generation

After training, the network has learned to predict the noise component that was used during the forward diffusion process. In image generation phase, the model starts from a pure sample from Gaussian noise as below:

$$z_T \sim \mathcal{N}(0,\, I)$$

At each timestep $t = T, T-1, \ldots, 1$, we feed the current noisy latent z_t, along with the conditioning vector $\tau(c)$ (which encodes the desired "class, domain" information), into the diffusion network. This diffusion network outputs

$$\hat{\epsilon}_\theta\big(z_t,\, t,\, \tau(c)\big)$$

which is the network's estimate of noise present in z_t. We then use this predicted noise to recover a "less-noisy" latent z_{t-1} given by

$$z_{t-1} \;=\; \sqrt{\frac{1}{\bar{\alpha}_t}}\Big(\, z_t \;-\; \frac{1-\bar{\alpha}_t}{\sqrt{1-\alpha_t}}\,\hat{\epsilon}_\theta\big(z_t,\, t,\, \tau(p)\big)\Big) \;+\; \sigma_t \eta \tag{10}$$

where $\bar{\alpha}_t = \frac{\alpha_t}{\alpha_{t-1}}$, and $\sigma_t = \sqrt{1-\bar{\alpha}_t}$. In Eqn.4 σ_t determines the fraction of the Gaussian noise sample $\eta \sim \mathcal{N}(0, I)$ that is injected at each timestep t of the reverse diffusion process to introduce controlled randomness during generation phase.

We repeat this denoising process recursively, from $t = T$ to $t = 0$, by predicting $\hat{\epsilon}_\theta\big(z_t, t, \tau(c)\big)$, and subtracting it from z_t to get z_{t-1} in each step. The latent image z_0 thus generated would belong to the domain and the class specified by the prompting conditioning $\tau(p)$ imposed on the model. Finally, z_0 is passed through a VAE decoder to obtain the generated image:

$$x_{\text{gen}} = \text{VAE-dec}(z_0).$$

In our approach, we enhance the generalization of the image generation capability of the diffusion model to generate images that belong to domains that lie in between the original training domains. To achieve this, we obtain the conditional embedding vector C as

$$C = \lambda \cdot \tau(p_1) + (1-\lambda) \cdot \tau(p_2) \tag{11}$$

to be used by the diffusion model, where $p_1 = (d_i, c_k)$ and $p_2 = (d_j, c_k)$ refer to the prompts corresponding to a pair of images from the domains d_i and d_j ($i \neq j$), respectively, but belonging to the same class c_k, and λ is set to 0.5. Eqn.5 controls the position of the generated image in the

feature space by generating an image that lies in between domains i and j, and the choice of the parameter $\lambda \in [0, 1]$ controls the proximity of the generated image to domain i or j in the feature space.

4.2.2 Formation of Intermediate Domains and Generalization of Classification

In the previous section, we proposed a scheme to generate clusters of synthetic images, in which each cluster contains a large number of images belonging to a region of the multidimensional feature space that lies between a pair of the original domains. In this section, we propose a scheme through which a new domain is formed for each of the clusters of the generated images, so that the new domain has a sufficient number of training images and is suitably positioned between the pair of the original training domains in question within the feature space.

We first extract the features of each of the images both in the original training domains and in the newly generated clusters using a feature extractor encoder (CPLIP [47]). The vectors representing the features of the images are plotted in the multidimensional feature space, as shown in Figure 4.1. Let $x_n^{d_i}$ denote the n-th patch ($n = 1, 2, \ldots, N_i$) from the original training domain d_i ($i = 1, 2, \ldots, I$), where N_i is the total number of patches in domain d_i, and I is the total number of original training domains. Let $E(x_n^{d_i})$ represent the feature vector of the patch $x_n^{d_i}$. We can calculate the centroidal feature using the features of all patches in domain d_i as

$$m_i = \frac{1}{N_i} \sum_{n=1}^{N_i} E(x_n^{d_i}). \tag{6}$$

The centroidal feature m_i, given by the above equation, can be considered to represent the features of all the patches in the domain d_i. Next, we form the intermediate domain $d_{i,j}$ for each pair (i, j), $i \neq j$, of the domains by including in it a suitable number of images $x_n^{d_{i,j}}$ from the cluster of images created in the neighborhood that lies between the domains d_i and d_j. For this purpose, we first compute a central feature for $d_{i,j}$ as

$$m_{i,j} = \frac{m_i + m_j}{2}. \tag{7}$$

With the central feature $m_{i,j}$ of the intermediate domain $d_{i,j}$, as given by the above equation, we now include in it K images from the cluster in question whose features are closest to the central feature $m_{i,j}$. In order to measure the closeness of the feature vector $E(x_n^{d_{i,j}}),\ n = 1, 2, \ldots, K$, of an image, to the central feature $m_{i,j}$, we use the Euclidean distance metric given by

$$d\big(E(x_n^{d_{i,j}}),\ m_{i,j}\big) = \Big\| E(x_n^{d_{i,j}}) - m_{i,j} \Big\| \tag{8}$$

It is what making the following remarks on our proposed scheme of forming the new intermediate domains. First of all, the central feature vector of a newly created domain $d_{i,j}$ is exactly in the middle of the centroids of the original domains d_i and d_j. We include in each of these new domains K images from the respective cluster, where K is a very large number (approximately 50,000 images). Since we generate a very large number (more than a million) of images for each cluster and choose K images from the cluster for each of the domains, it can be expected that the central feature $m_{i,j}$ is very close to the centroid of the domain $d_{i,j}$. Hence, the central feature $m_{i,j}$ can be expected to represent the features of all the images included in the domain $d_{i,j}$.

Finally, an enlarged training set that combines all the images from the newly created domains and those from the existing training domains is used for training a classifier model. A training set so formed can be expected to expand the generalization capability of the classifier in dealing with the classification of samples with distributions shifted from that of the original training domains.

4.3 Experimental Results

Following the methodology described in the previous section, this section presents a comprehensive evaluation of the proposed PathoGen framework for enhancement of generalization capability of the classifier models in histopathology image classification. We begin with the introduction of the Camelyon17 WILDS dataset [17] that we use in our experiments, followed by a description of the software and hardware platforms used for implementing the proposed scheme. We then conduct a number of experiments in order to provide a comprehensive evaluation of our PathoGen scheme. An ablation study is also carried out to further validate the effectiveness of various ideas employed in developing the proposed scheme. Finally, we compare the performance of the proposed PathoGen

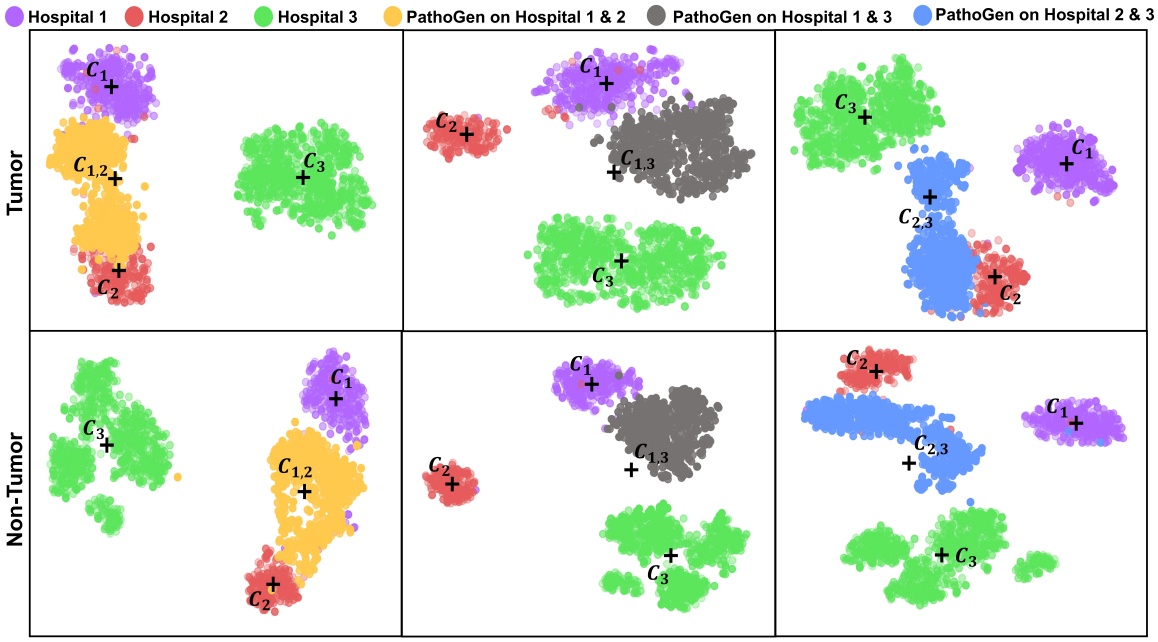

Figure 4.3: T-SNE Visualization of Tumor and Non-Tumor Clusters Across Original and Newly Generated Training Sets. The Plot Shows Distinct Clusters for the Tumor and Non-tumor Categories Across Three Hospitals (D_1: Purple, D_2: Red, D_3: Green), with PathoGen-generated Synthetic Samples Interpolating between them (Yellow, Gray, Blue).

scheme with that of existing state-of-the-art schemes for enhancing the generalization capability of classifier models.

4.3.1 Dataset

In this study, we use the Camelyon17 WILDS dataset [17], which was previously introduced and described in detail in Chapter 3. We follow the same domain partitioning protocol used in the previous chapter, where training is performed on domains D_1 to D_3, validation on D_4, and testing on the unseen domain D_5. This consistent setup ensures a fair comparison across methods and enables a rigorous evaluation of the generalization capability of the trained classifiers.

4.3.2 Visualization of Original and Intermediate Domains

To assess the effectiveness of the proposed PathoGen framework qualitatively in synthesizing intermediate domains, we present in Fig. 4.3 the t-SNE visualizations of the feature vectors of the images in the original domains D_1, D_2, D_3 of the Camelyon17 WILDS dataset, and those of the

images in the newly created domains $D_{1,2}$, $D_{2,3}$, $D_{3,1}$. From the structure and relative positions of the feature points of the images, it is clearly seen that when the proposed method of domain formation is applied to the images of the original domains of a real-life dataset, it is successful in generating images for the new intermediate domains.

4.3.3 Implementation Details

We implement our proposed method using Python, with the PyTorch deep learning framework. All the experiments have been conducted on a system equipped with four NVIDIA A100 GPUs. Our diffusion model is initialized using pretrained weights from Stable Diffusion version 1.5, which serve as the starting point for training. To train our stable diffusion model further, we have used a learning rate of 1e-4 with the LambdaLinear learning rate scheduler. The batch size used is set to 96, distributed across 24 samples per GPU. We use 16-bit floating point precision (FP16) during training to reduce memory usage and speed up computation without sacrificing accuracy. Training was conducted for a maximum of 10,000 steps, where each step represents one batch of training data passed through the model. During the image generation, model goes through 1000 timesteps to gradually remove the noise from the input to produce a final synthetic image. We employed the DDPM sampler [25] for the diffusion process. We use image patches of size 256×256×3 for the autoencoder and reduce it to the size 32×32×4 for its latent representation for a more efficient training and image generation. The model so trained is then used to generates 1 million synthetic images for each of the two class, tumor, and non-tumor. As for the extraction of the features providing a feature vector for each of the original and synthetically generated images, we use CPLIP [47] vision encoder model. In our experiments, we set the number of selected synthetic samples per class to be $K = 50$k from the 1 million generated images. This choice for the value of K is based on an empirical study in which we plot the accuracy of the classification as a function of K, the number of images selected for each of the newly created domains, as shown in Fig. 4.4. It is seen from this figure that the classification accuracy is maximized when K is selected to be 50k. It is to be noted that $K = 50$k happens to be approximately the average number of images in the original domains. In order to demonstrate the effectiveness of our proposed scheme for domain generalization, we use ResNet-50 [48] and the original Vision Transformer (ViT-B/16) [49] as backbone classifiers.

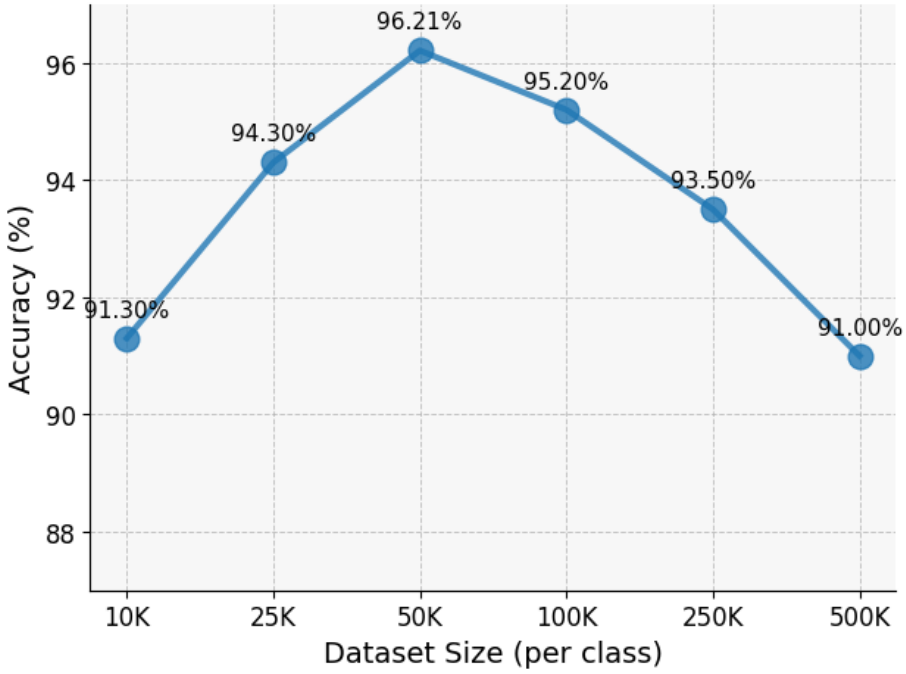

Figure 4.4: Effect of Sample Size K on the Performance.

To have a uniform performance evaluation of the classifier across different approaches for domain generalization, we use the DomainBed framework [50] for the training and testing of the classifier.

4.3.4 Performance Evaluation

To evaluate the generalization performance of our proposed *PathoGen* framework, we compare the classification accuracy of the two classifiers, *ResNet-50* and *ViT*, each trained on two different training configurations, Config 1 and Config 2. In configuration 1, the classifiers are trained using a training set consisting of images from three original training domains, namely D_1, D_2, and D_3. In configuration 2, we extend the configuration 1 training set by adding to it the synthetic images generated from the intermediate domains, $D_{1,2}$, $D_{2,3}$, and $D_{3,1}$. Each of the trained models of a classifier is then tested on two test sets: Test Set 1 and Test Set 2. Test Set 1 includes held-out samples drawn from the source domains, namely D_1, D_2, and D_3, allowing us to assess performance of a classifier in an in-domain setting. Test Set 2, on the other hand, contains samples from domains not included in D_1, D_2, and D_3. In our experiments, the images in Test Set 2 are selected from hospital centre H_5, that is, the images from domain D_5. Thus, the two trained models are

Table 4.1: Comparison of Classification Accuracy (%) of Models Trained with Different Training Configurations and Tested on Test Set 1 and Test Set 2.

Training Configuration	**ResNet**		**ViT**	
	Test Set 1	**Test Set 2**	**Test Set 1**	**Test Set 2**
Config 1 (D_1, D_2, D_3)	97.0	70.3	97.2	71.1
Config 2 (D_1, D_2, D_3, $D_{1,2}$, $D_{2,3}$, $D_{3,1}$ **(Proposed)**)	**98.2**	**96.2**	**98.4**	**97.4**

evaluated using Test Set 2 for their generalization capability under the real-world challenge of out-of-distribution (OOD) shift of the test samples. The results, in terms of classification accuracy of the two classifiers trained with the two different training configurations and tested on the two test sets, are presented in Table 4.1. It is seen from this table that for either of the two classifiers, when the proposed scheme for domain generalization is not used (the results in the first row of Table 4.1), we observe a significant drop in the performance under domain shift in the test samples. For example, the performance of the ResNet classifier degrades from an accuracy of 97.0% to 70.30%. On the other hand, when the proposed scheme is applied to the classifiers (the second row of Table 4.1), the classifiers become highly robust to domain shifts in the test samples. For example, for the ViT transformer, the classification accuracy of 98.48% drops only by one percentage point to 97.35%. From the results on Test Set 2 in Table 4.1, it is especially noteworthy that the proposed scheme has a very significant impact on enhancing the classification accuracy, in that it increases the classification accuracy from 70.30% to 96.21% when the ResNet classifier is used, and from 71.12% to 97.35% when the ViT classifier is used.

4.3.5 Comparison with Existing Methods

In this section, we compare the effectiveness of our proposed domain generalization scheme, PathoGen, with that of a number of state-of-the-art schemes. For this purpose, we compare the classification accuracy obtained by applying the various domain generalization schemes on the same dataset, namely Camelyon17-WILDS, in which the domains D_1, D_2, and D_3 are specified to be used for training the classifier, and D_4 and D_5, specifically for validation and testing. We have selected two classifiers, ResNet-50 [48] and ViT [49], the same two classifiers that have been used by the various other schemes, for this comparison. The classification accuracy is reported in two

separate tables, one for each of the two classifiers. Ideally, it would be desirable to include all the proposed schemes in both tables for a comprehensive comparison. However, this has not been possible due to the fact that not all the schemes have reported their accuracy performance on both classifiers. To show the effectiveness of the domain generalization methods in the literature, researchers have compared the performance accuracy of their methods with that of the 2021 baseline non domain generalization scheme, [50], for multi-domain image classification. In this scheme, the classifier used undergoes an in-domain training, that is, it is trained using the images from D_1, D_2, D_3 and tested on D_5. When the performance accuracy of a proposed scheme is compared with that of ERM, the difference in the two performances indicates as to how effective is the proposed scheme in achieving the domain generalization.

Table 4.2 provides the accuracy results of our proposed PathoGen scheme and with that of 13 other domain generalization schemes, along with that of ERM. It is seen from this table that all the methods proposed after 2021 have succeeded in handling the domain shift in the test image by providing performance accuracy higher than that of but with varying amounts. It is noted that PathoGen provides the highest classification accuracy of 96.2%. This classification accuracy of the proposed scheme is 1.8% and 2.2% higher than that of the second-best scheme, PathoWAve [27], and the third best scheme TTI2I [51], respectively. In addition to improvements in accuracy, it is important to note that the second- and third-best methods also exhibit drawbacks not shared by our proposed approach.

It is worth noting that, the second-best method, PathoWAve [27]—the scheme proposed in Chapter 3- improves diversity using general and pathology-specific augmentations, but these operate only within source domains and do not generate intermediate-domain samples representative of unseen domains. Consequently, the feature space covered during training remains discontinuous across domains, limiting the model's ability to generalize to truly novel acquisition conditions. To address this limitation, the proposed PathoGen framework explicitly expands the domain coverage by generating synthetic samples that populate the previously uncovered intermediate regions of the feature space, thereby creating a more continuous and domain-invariant training manifold. Also in comparison to the third best method [51], which relies on style transfer using generative models at test time thus introducing significant inference computational complexity, the proposed scheme does

Table 4.2: Comparative Performance of Various Domain Generalization Methods using ResNet 50 Classifier Trained with the Training Configuration Config 3 and Tested on Test Set 2. The Best, Second-Best, and the Third-Best Results are Represented in Red, Blue, and Green, Respectively.

Method	**Accuracy %**
ERM (2021) [50]	70.3
CORAL (2016) [52]	59.5
IRM (2019) [53]	64.2
Group DRO (2019) [54]	68.4
DomainMix (2020) [55]	69.7
MMLD (2020) [56]	70.2
FISH (2021) [57]	74.7
V_REx (2021) [58]	71.5
STRAP (2021) [59]	93.7
LISA (2022) [14]	77.1
StarGanV2 (2022) [51]	76.4
TTI2I (2022) [51]	94.0
FuseStyle (2023) [60]	90.5
PathoWAve (2024) [27]	94.4
PathoGen (Proposed)	**96.2**

not introduce any test time complexity overhead.

Table 4.3 depicts the classification performance of various domain generalization schemes using the ViT classifier. Our PathoGen scheme shows the best performance with the classification accuracy of 97.4% which is 1.2% higher than the accuracy obtained with the ResNet classifier. The schemes providing the second and third best performance are, respectively, PathoWAve [27], and EPVT [16], with the accuracy that are, respectively, 2.5% and 11% lower than that provided by the proposed scheme. Another important distinction between the proposed PathoGen scheme and the EPVT scheme that must be noted is that the latter scheme is designed only for vision transformer as a classifier, whereas the former is a general scheme that can be used by any type of classifier model.

Table 4.3: Comparative Performance of Various Domain Generalization Methods using ViT Classifier Trained with the Training Configuration Config 3 and Tested on Test Set 2. The Best, Second Best, and Third Best Results are Represented in Red, Blue, and Green, Respectively.

Method	**Accuracy %**
ERM (2021) [50]	73.1
CORAL (2016) [52]	71.8
DANN (2016) [61]	83.5
IRM (2019) [53]	75.0
PLDG (2024) [15]	84.3
EPVT (2024) [16]	86.4
PathoWAve (2024) [27]	94.9
PathoGen (Proposed)	**97.4**

4.3.6 Ablation Studies

Impact of In-Domain and Cross-Domain Data Generation: In this section, we evaluate the impact of synthetic domain generation within individual training domains and across multiple domains using PathoGen's interpolation strategy. The results are summarized in Table 4.4, where we compare two methods with the baseline (ERM). The method with Configuration 3 training includes the original images belonging to D_1, D_2,and D_3 plus the synthetic images within D_1, D_2, and D_3. For ResNet-50 classifier, ERM, which serves as a baseline without using any extra synthetic images, achieves a test accuracy of 70.3%, reflecting the model's performance when trained exclusively on the original training data. In-domain image generation, where the diffusion model produces additional samples within each training domain but without interpolation across domains, improves generalization, reaching 75.8% accuracy. This demonstrates that exposure to additional intra-domain variability helps the model adapt to new data. Finally, PathoGen's domain-interpolated generation significantly boosts performance, achieving 96.2% test accuracy. Similarly, for the ViT classifier, we observe the same trend. The baseline achieves a test accuracy of 71.1%, in-domain generation improves it to 76.3%, and PathoGen further increases performance to 97.4%. These results

Table 4.4: Comparison of in-domain data generation with our proposed PathoGen generation (cross-domain) on Test Set 2, which Contains Samples from Unseen Domains. Results are Reported for both ResNet-50 and ViT Classifiers.

Method	**ResNet-50** Accuracy (%)	**ViT** Accuracy (%)
ERM (Baseline) [50]	70.3	71.1
Method with Config 3 training	75.8	76.3
PathoGen (ours)	**96.2**	**97.4**

highlights the advantage of domain interpolation strategy, where generating samples that lie between training domains in feature space better captures variations seen in real-world histopathology datasets. While in-domain generation provides some benefit, PathoGen's domain-interpolated synthetic data significantly enhances generalization, making it far more effective in handling domain shift and previously unseen samples.

Impact of Placement of the Intermediate Domain in the Feature Space: Recall that in our proposed scheme the intermediate domains were placed with a center which is the midpoint between the centers of the pairs of domains in question, that is, the center of the domain $D_{i,j}$ is chosen to be the midpoint of the centers of the domains of D_i and D_j in the feature space (a domain pair based center selection). In this ablation study, we investigate the impact on the accuracy performance of the classifiers when the center of $D_{i,j}$ is selected to be the center of all the images generated in the regions between the original domains D_i and D_j, that is, we have cluster-based centers for the new domains. With the center of the domain $D_{i,j}$ so chosen, the size of the domain $D_{i,j}$ is then obtained in the same way as that in the proposed domain generation scheme. The results of this ablation study are given in Table 4.5, in which the first row gives the classification accuracy of the two classifiers when the centers of the intermediate domains are cluster-based, whereas the second row provides the results when the centers of the intermediate domains are chosen to be pair-based (the proposed scheme). A comparison of the results in the two rows of the table shows that the proposed scheme of the placement of the generated intermediate domains is more effective in domain generalization. This is in view of the fact that the images selected in a new domain have a distribution which lies

Table 4.5: Comparison of Different Placement of the Intermediate Domain in the Feature Space.

Method	**ResNet-50** Accuracy (%)	**ViT** Accuracy (%)
Cluster-based center	95.7	96.8
PathoGen (Ours)	**96.2**	**97.4**

more between the distributions of the images in the pair of domains in question.

4.4 Comparison between PathoWAve and PathoGen

Both PathoWAve and PathoGen are designed to improve the domain generalization capability of histopathology classifiers by mitigating the adverse effects of domain shift across medical centers. However, they address this challenge from two fundamentally different perspectives. PathoWAve, presented in Chapter 3, focuses on robust optimization. It employs a multi-trajectory training strategy combined with weight averaging and diverse augmentations to guide the model toward flatter minima in the loss landscape and to increase robustness to local variations such as staining or imaging differences. These augmentations, however, are confined to the neighborhood of the original source domains and thus improve generalization only within limited, locally perturbed regions of the feature space.

In contrast, PathoGen, proposed in Chapter 4, approaches the problem from the data-space perspective. Rather than relying solely on local perturbations of existing data using augmentation techniques, PathoGen explicitly enhances the continuity of the feature space by creating intermediate domains between pairs of original domains. This is achieved by leveraging a conditional Stable Diffusion model to generate realistic synthetic images that populate the previously uncovered regions between domains. The resulting expanded training set provides a more continuous and better-connected representation of the data manifold, enabling the classifier to generalize more effectively to unseen hospitals.

4.5 Summary

In the feature space, domain shifts of histopathology images are generally continuous within a bounded region. Hence, if the feature space corresponding to a given training set consisting of a certain number of domains can be made more continuous by adding to it additional domains, then the training of a classifier with such a modified training set can be expected to handle the domain shift problem better than a classifier trained with the original training set. Accordingly, in this paper we have proposed a novel domain generalization method, called PathoGen, in which the original training set is modified to include new training domains that are in between each pair of the original domains, thereby making the corresponding feature space more continuous and training the classifier with this modified training set. A conditional stable diffusion model has been leveraged for generating synthetic images for the newly formed intermediate domains. The proposed domain generalization technique has been tested on the Camelyon17 WILDS dataset, a dataset widely used to validate existing domain generalization methods. The performance of the proposed scheme has been shown to sig- nificantly outperform existing domain generalization schemes.

Chapter 5

Conclusion and Scope for future investigation

5.1 Concluding the marks

This thesis presented two novel methodologies—PathoWAve and PathoGen—aimed at addressing domain generalization challenges in histopathology image analysis. Both methods contribute significantly to improving the generalization of models across multiple unseen domains, particularly in the face of domain shifts caused by variations in imaging equipment, staining techniques, and institutional differences for histopathology image classification without increasing test time complexity and having access to the test data during training.

In Chapter 3, we introduced PathoWAve, a deep learning-based weight averaging training startegy designed to improve domain generalization in histopathology image classification task. PathoWAve employs a multi-trajectory training approach combined with cyclical weight averaging and a comprehensive set of general and histopathology-specific data augmentations. The method enables models to generalize effectively across multiple unseen domains without requiring access to test-time data.

Chapter 4 introduced PathoGen, an innovative framework that enhances the continuity of the domains in the feature space of the original training set. Specifically, in this proposed scheme the continuity is enhanced by creating a new intermediate domain in the training set between each pair

of its original domains. The images for the newly created intermediate domains are generated by a conditional Stable diffusion model. Finally, the training set thus modified is used to train a classifier model for the classification of the histopathology test images characterized by domain shifts.

Our experiments on the Camelyon17 WILDS dataset demonstrated that both proposed schemes, PathoWAve and PathoGen, outperforms state-of-the-art domain generalization methods by achieving higher accuracy and robustness, making it a valuable contribution to domain generalization research.

5.2 Scope for future work

Future research could extend the current pairwise domain interpolation approach to go through more cycles of generating the new intermediate domains, and generate intermediate domains more and more in between each pair of the domains after the first cycle to increase the continuity of the feature space more, allowing the model to handle domain shift more effectively.

References

[1] P.P. Shinde and S. Shah, "A review of machine learning and deep learning applications," in *2018 Fourth international conference on computing communication control and automation (ICCUBEA)*. IEEE, 2018, pp. 1–6.

[2] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *Ieee Access*, vol. 6, pp. 9375–9389, 2017.

[3] K. Stacke, G. Eilertsen, J. Unger, and C. Lundström, "Measuring domain shift for deep learning in histopathology," *IEEE journal of biomedical and health informatics*, vol. 25, no. 2, pp. 325–336, 2020.

[4] Y. Luo, L. Zheng, T. Guan, J. Yu, and Y. Yang, "Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2507–2516.

[5] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C.C. Loy, "Domain generalization: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4396–4415, 2022.

[6] L. Chen, Y. Zhang, Y. Song, Y. Shan, and L. Liu, "Improved test-time adaptation for domain generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 24 172–24 182.

[7] J.S. Yoon, K. Oh, Y. Shin, M.A. Mazurowski, and H.I. Suk, "Domain generalization for medical image analysis: A survey," *arXiv preprint arXiv:2310.08598*, 2023.

[8] P. Izmailov, D. Podoprikhin, T. Garipov, D. Vetrov, and A.G. Wilson, "Averaging weights leads to wider optima and better generalization," *arXiv preprint arXiv:1803.05407*, 2018.

[9] J. Cha, S. Chun, K. Lee, H.C. Cho, S. Park, Y. Lee, and S. Park, "Swad: Domain generalization by seeking flat minima," *Advances in Neural Information Processing Systems*, vol. 34, pp. 22 405–22 418, 2021.

[10] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 443–450.

[11] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant risk minimization," *arXiv preprint arXiv:1907.02893*, 2019.

[12] S. Sagawa, P.W. Koh, T.B. Hashimoto, and P. Liang, "Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization," *arXiv preprint arXiv:1911.08731*, 2019.

[13] Y. Shi, J. Seely, P.H. Torr, N. Siddharth, A. Hannun, N. Usunier, and G. Synnaeve, "Gradient matching for domain generalization," *arXiv preprint arXiv:2104.09937*, 2021.

[14] H. Yao, Y. Wang, S. Li, L. Zhang, W. Liang, J. Zou, and C. Finn, "Improving out-of-distribution robustness via selective augmentation," in *International Conference on Machine Learning*. PMLR, 2022, pp. 25 407–25 437.

[15] S. Yan, Z. Yu, C. Liu, L. Ju, D. Mahapatra, B. Betz-Stablein, V. Mar, M. Janda, P. Soyer, and Z. Ge, "Prompt-driven latent domain generalization for medical image classification," *IEEE Transactions on Medical Imaging*, 2024.

[16] S. Yan, C. Liu, Z. Yu, L. Ju, D. Mahapatra, V. Mar, M. Janda, P. Soyer, and Z. Ge, "Epvt: Environment-aware prompt vision transformer for domain generalization in skin lesion recognition," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 249–259.

[17] P.W. Koh, S. Sagawa, H. Marklund, S.M. Xie, M. Zhang, A. Balsubramani, W. Hu, M. Yasunaga, R.L. Phillips, I. Gao *et al.*, "Wilds: A benchmark of in-the-wild distribution shifts," in *International conference on machine learning*. PMLR, 2021, pp. 5637–5664.

[18] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)*, 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.

[22] J.L. Ba, J.R. Kiros, and G.E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[23] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[24] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[25] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *arXiv preprint arXiv:2006.11239*, 2020.

[26] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.

[27] P. Sotoudeh Sharifi, M.O. Ahmad, and M.N.S. Swamy, "Pathowave: A deep learning-based weight averaging method for improving domain generalization in histopathology images," in *2024 IEEE 67th International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2024, pp. 975–979.

[28] E.D. Cubuk, B. Zoph, J. Shlens, and Q.V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 702–703.

[29] D. Tellez, M. Balkenhol, I. Otte-Höller, R. van de Loo, R. Vogels, P. Bult, C. Wauters, W. Vreuls, S. Mol, N. Karssemeijer *et al.*, "Whole-slide mitosis detection in h&e breast histology using phh3 as a reference to train distilled stain-invariant convolutional networks," *IEEE transactions on medical imaging*, vol. 37, no. 9, pp. 2126–2136, 2018.

[30] M. Xu, J. Zhang, B. Ni, T. Li, C. Wang, Q. Tian, and W. Zhang, "Adversarial domain adaptation with domain mixup," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 04, 2020, pp. 6502–6509.

[31] T. Matsuura and T. Harada, "Domain generalization using a mixture of multiple latent domains," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 11 749–11 756.

[32] D. Krueger, E. Caballero, J.H. Jacobsen, A. Zhang, J. Binas, D. Zhang, R. Le Priol, and A. Courville, "Out-of-distribution generalization via risk extrapolation (rex)," in *International conference on machine learning*. PMLR, 2021, pp. 5815–5826.

[33] K. Ahuja, E. Caballero, D. Zhang, J.C. Gagnon-Audet, Y. Bengio, I. Mitliagkas, and I. Rish, "Invariance principle meets information bottleneck for out-of-distribution generalization," *Advances in Neural Information Processing Systems*, vol. 34, pp. 3438–3450, 2021.

[34] H. Yao, Y. Wang, S. Li, L. Zhang, W. Liang, J. Zou, and C. Finn, "Improving out-of-distribution robustness via selective augmentation," in *International Conference on Machine Learning*. PMLR, 2022, pp. 25 407–25 437.

[35] V. Khamankar, S. Bera, S. Bhattacharya, D. Sen, and P.K. Biswas, "Histopathological image analysis with style-augmented feature domain mixing for improved generalization," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 285–294.

[36] R. Yamashita, J. Long, S. Banda, J. Shen, and D.L. Rubin, "Learning domain-agnostic visual representation for computational pathology using medically-irrelevant style transfer augmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3945–3954, 2021.

[37] M. Scalbert, M. Vakalopoulou, and F. Couzinié-Devy, "Test-time image-to-image translation ensembling improves out-of-distribution generalization in histopathology," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 120–129.

[38] S. Yan, Z. Yu, C. Liu, L. Ju, D. Mahapatra, B. Betz-Stablein, V. Mar, M. Janda, P. Soyer, and Z. Ge, "Prompt-driven latent domain generalization for medical image classification," *IEEE Transactions on Medical Imaging*, 2024.

[39] D. Kim, Y. Yoo, S. Park, J. Kim, and J. Lee, "Selfreg: Self-supervised contrastive regularization for domain generalization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9619–9628.

[40] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of machine learning research*, vol. 17, no. 59, pp. 1–35, 2016.

[41] S. Yan, C. Liu, Z. Yu, L. Ju, D. Mahapatra, V. Mar, M. Janda, P. Soyer, and Z. Ge, "Epvt: Environment-aware prompt vision transformer for domain generalization in skin lesion recognition," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 249–259.

[42] P. Sotoudeh Sharifi, M.O. Ahmad, and M.N.S. Swamy, "PathoGen: A generative diffusion-based domain generalization scheme for histopathology image classification," 2025, under review for a journal publication.

[43] F. Aeffner, M.D. Zarella, N. Buchbinder, M.M. Bui, M.R. Goodman, D.J. Hartman, G.M. Lujan, M.A. Molani, A.V. Parwani, K. Lillard, and M.A. Berman, "Introduction to digital image analysis in whole slide imaging: A white paper from the digital pathology association," *Journal of Pathology Informatics*, vol. 10, no. 1, p. 9, 2019.

[44] S.R. Duenweg, A.M. Dahm, C.S. McGary, S. Chen, K.A. Pearlstein, A.M. Nelson, G. Venkataraman, O. Ardon, R.M. Levenson, E.S. Reisenbichler, and M.G. Hanna, "Whole slide imaging scanner differences influence quantitative image analysis in histopathology," *Archives of Pathology & Laboratory Medicine*, vol. 147, no. 9, pp. 1104–1115, 2023.

[45] F. Wilm, M. Aubreville *et al.*, "Mind the gap: Scanner-induced domain shifts pose challenges for representation learning in histopathology," *arXiv preprint arXiv:2211.16141*, 2022.

[46] A.Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," *arXiv preprint arXiv:2102.09672*, 2021.

[47] S. Javed, A. Mahmood, I.I. Ganapathi, F.A. Dharejo, N. Werghi, and M. Bennamoun, "Cplip: Zero-shot learning for histopathology with comprehensive vision-language alignment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 11 450–11 459.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[49] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[50] I. Gulrajani and D. Lopez-Paz, "In search of lost domain generalization," *arXiv preprint arXiv:2007.01434*, 2020.

[51] M. Scalbert, M. Vakalopoulou, and F. Couzinié-Devy, "Test-time image-to-image translation ensembling improves out-of-distribution generalization in histopathology," in *International*

Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2022, pp. 120–129.

[52] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Computer vision–ECCV 2016 workshops: Amsterdam, the Netherlands, October 8-10 and 15-16, 2016, proceedings, part III 14*. Springer, 2016, pp. 443–450.

[53] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant risk minimization," *arXiv preprint arXiv:1907.02893*, 2019.

[54] S. Sagawa, P.W. Koh, T.B. Hashimoto, and P. Liang, "Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization," *arXiv preprint arXiv:1911.08731*, 2019.

[55] M. Xu, J. Zhang, B. Ni, T. Li, C. Wang, Q. Tian, and W. Zhang, "Adversarial domain adaptation with domain mixup," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 04, 2020, pp. 6502–6509.

[56] T. Matsuura and T. Harada, "Domain generalization using a mixture of multiple latent domains," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 11 749–11 756.

[57] Y. Shi, J. Seely, P.H. Torr, N. Siddharth, A. Hannun, N. Usunier, and G. Synnaeve, "Gradient matching for domain generalization," *arXiv preprint arXiv:2104.09937*, 2021.

[58] D. Krueger, E. Caballero, J.H. Jacobsen, A. Zhang, J. Binas, D. Zhang, R. Le Priol, and A. Courville, "Out-of-distribution generalization via risk extrapolation (rex)," in *International conference on machine learning*. PMLR, 2021, pp. 5815–5826.

[59] R. Yamashita, J. Long, S. Banda, J. Shen, and D.L. Rubin, "Learning domain-agnostic visual representation for computational pathology using medically-irrelevant style transfer augmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3945–3954, 2021.

[60] V. Khamankar, S. Bera, S. Bhattacharya, D. Sen, and P.K. Biswas, "Histopathological image analysis with style-augmented feature domain mixing for improved generalization," in

International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2023, pp. 285–294.

[61] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of machine learning research*, vol. 17, no. 59, pp. 1–35, 2016.