Disentangling Distribution Shift

The key to robust computer vision models

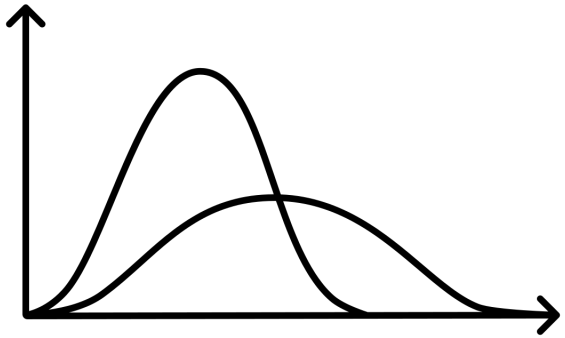
Heather Couture

March 6, 2025
12 pm EST

30 minutes + Q&A

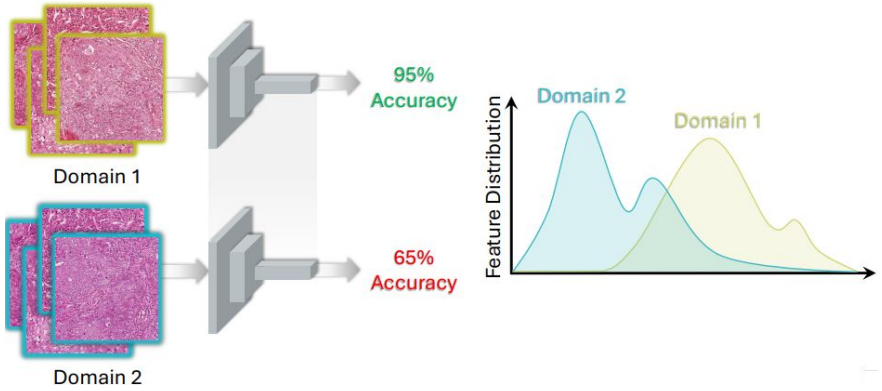
Pixel Scientia LABS

Definition of Distribution Shift



Source !=
Target

Why Distribution Shift Matters



Source: Zamanitajeddin, Benchmarking Domain Generalization Algorithms in Computational Pathology, 2024

- Decreased accuracy
- Potential for biased or unfair outcomes
- Unexpected behavior in real-world scenarios
- Safety concerns in critical applications

Goals of this webinar

What distribution shift is

Different types of distribution shift

How to identify distribution shift

How to address distribution shift

Who am I?

- Heather Couture
- PhD in Computer Science from University of North Carolina
- Contributor to Scientific American, The Pathologist, IEEE Spectrum
- Newsletter and podcast

Computer Vision Insights
by Pixel Scientia Labs

- Computer vision consultant

OWKIN DECIPHEX ULTIVUE WattTime Leica MICROSYSTEMS QRITIVE

Causes of Distribution Shift

Technical shifts:

- Changes in image resolution
- Variations in lighting conditions
- Different camera angles or perspectives
- Different imaging devices
- Differences in sample preparation
  - Staining techniques
  - Frozen vs. FFPE tissue sections
- Artifacts
- Different annotators

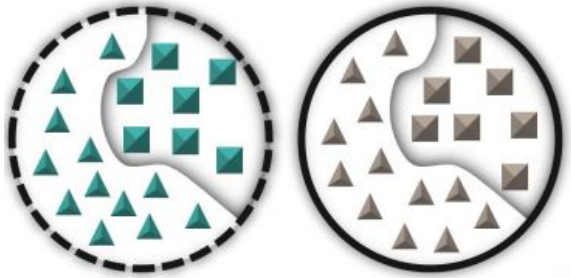
Biological or geological shifts:

- Seasonal variations:
  - Wet vs. dry season
  - Spring vs. fall
- Different geographic regions
- Landscape alterations caused by natural disasters
- Biological variants
- Demographics: age, sex, race
- Location on the body of tissue sample

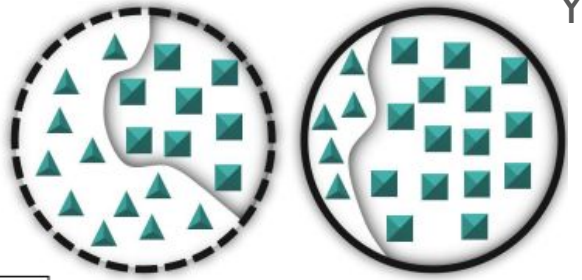
Types of Distribution Shifts

S = Source
T = Target
X = features
Y = labels

$P(X_S) \neq P(X_T)$

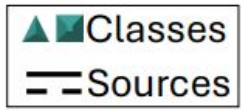


Covariate Shift

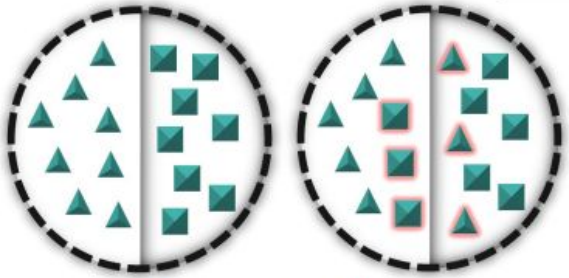


Prior Shift

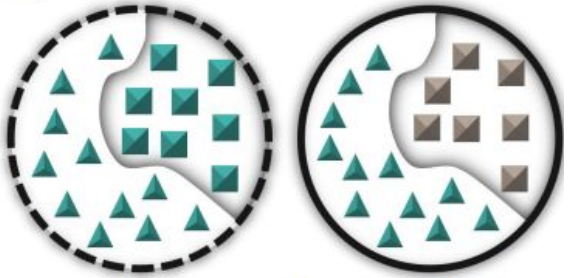
$P(Y_S) \neq P(Y_T)$



$P(Y_S|X_S) \neq P(Y_T|X_T)$



Posterior Shift

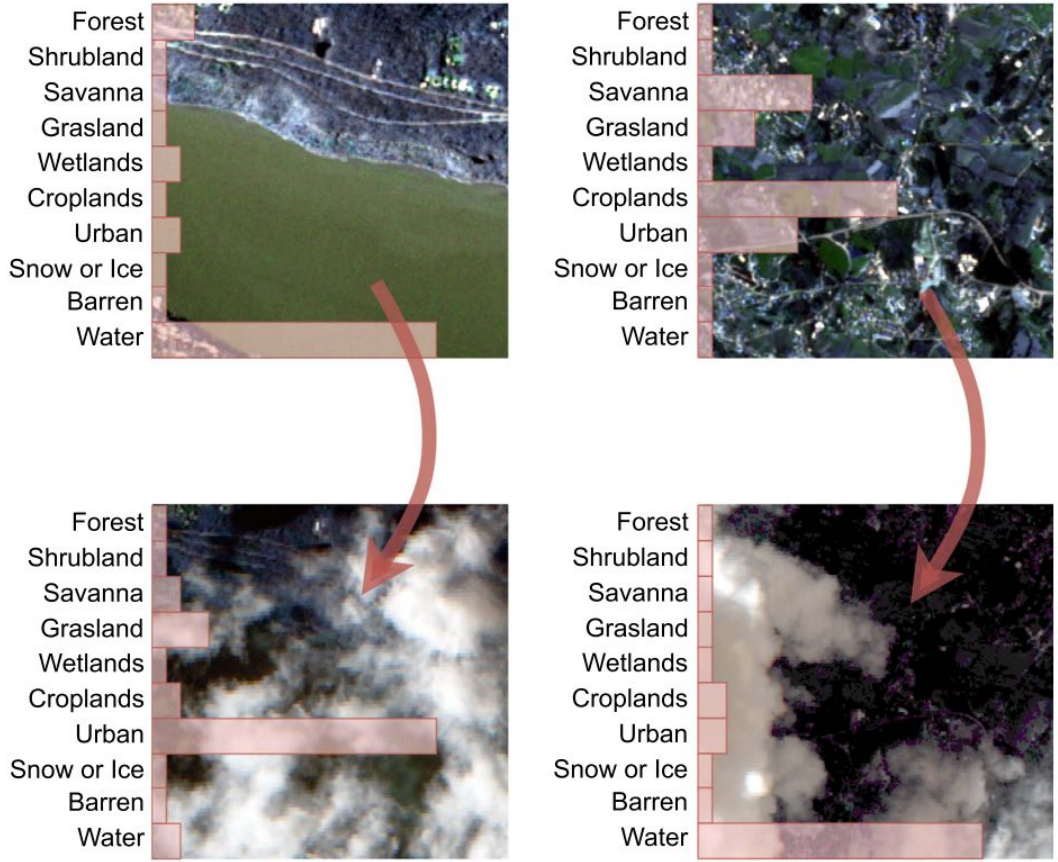


Class-conditional Shift

$P(X_S|Y_S=y) \neq P(Y_T|X_T=y)$

Source: Zamanitajeddin, Benchmarking Domain Generalization Algorithms in Computational Pathology, 2024

Example: Clouds

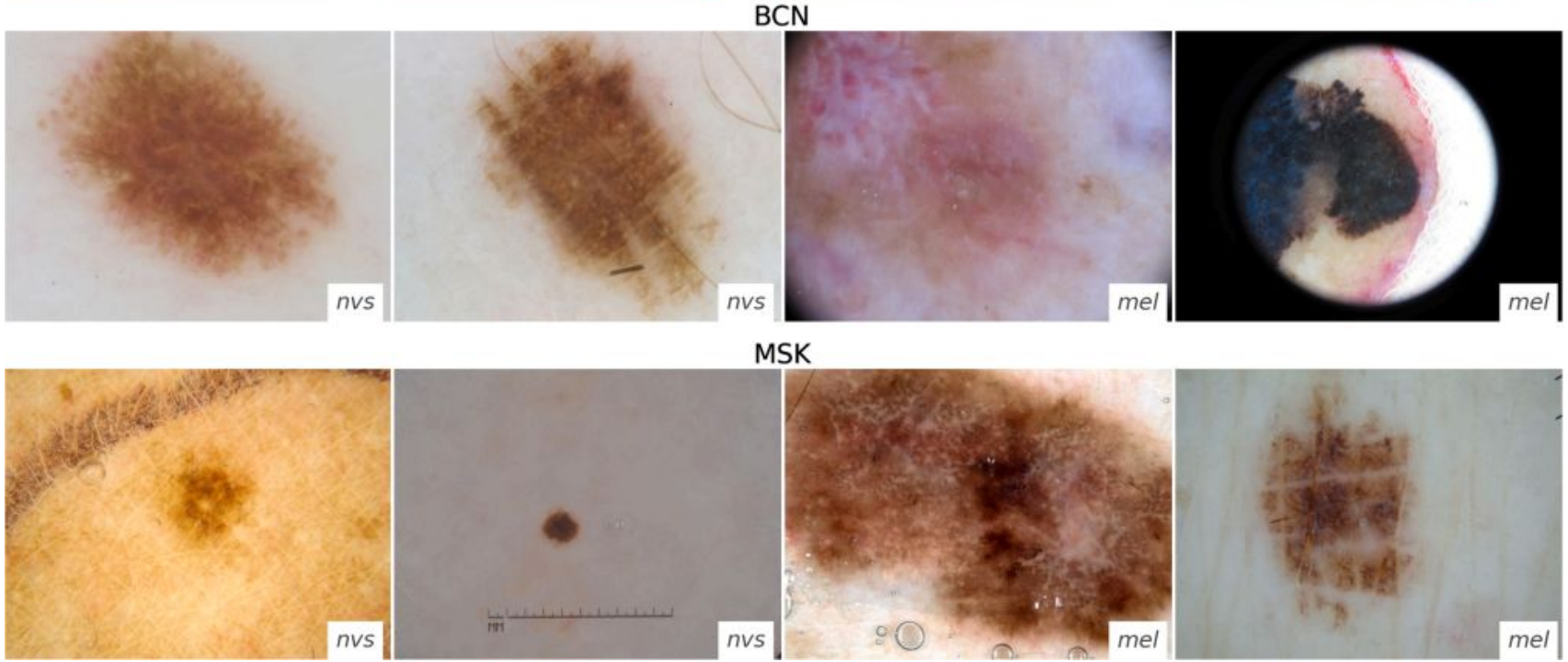


What type of shift is this?

Source: Gawlikowski, Explaining the Effects of Clouds on Remote Sensing Scene Classification, 2022

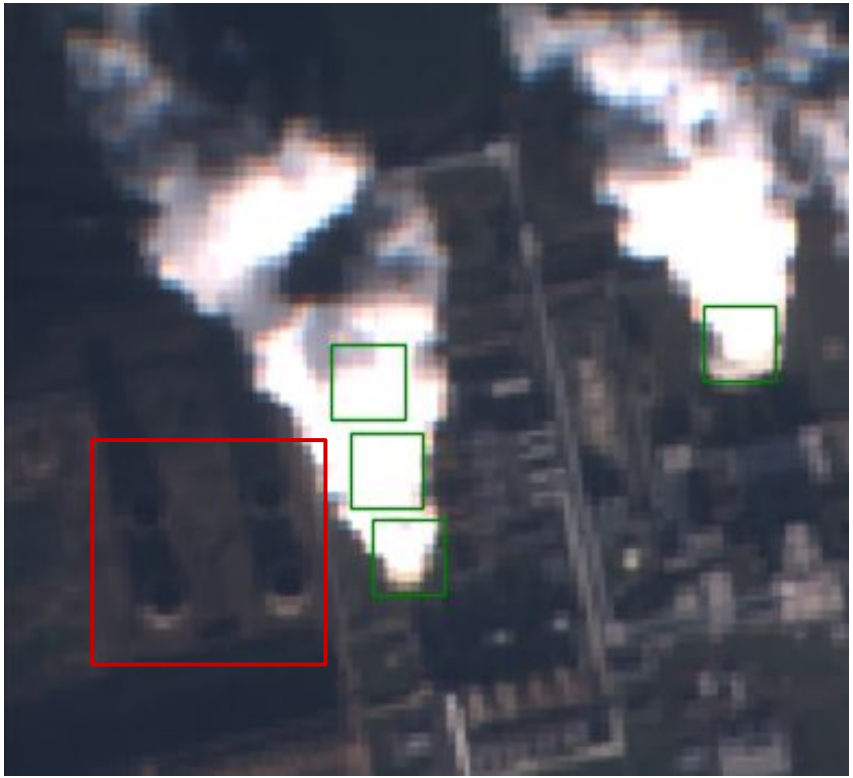
Example: Hospitals

What type of shift is this?



Source: Fogelberg, Domain shifts in dermoscopic skin cancer datasets: Evaluation of essential limitations for clinical translation, 2023

Example: Power Plant Utilization



4 cooling towers decommissioned

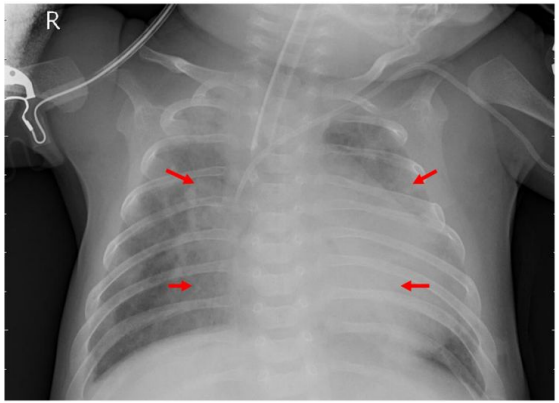
Google Maps view of plant

Model predicts 45% utilization when plant is operating at 95%

What type of shift is this?

Example: Chest X-Rays

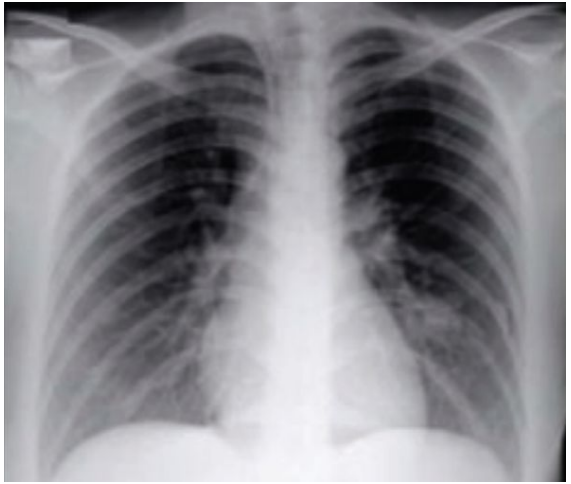
Healthy



Pneumonia

Source: Kundu, Pneumonia detection in chest X-ray images using an ensemble of deep learning models, 2021

Covid-19 Pneumonia



Source: Al-Waisy, COVID-DeepNet: Hybrid Multimodal Deep Learning System for Improving COVID-19 Pneumonia Detection in Chest X-ray Images, 2021

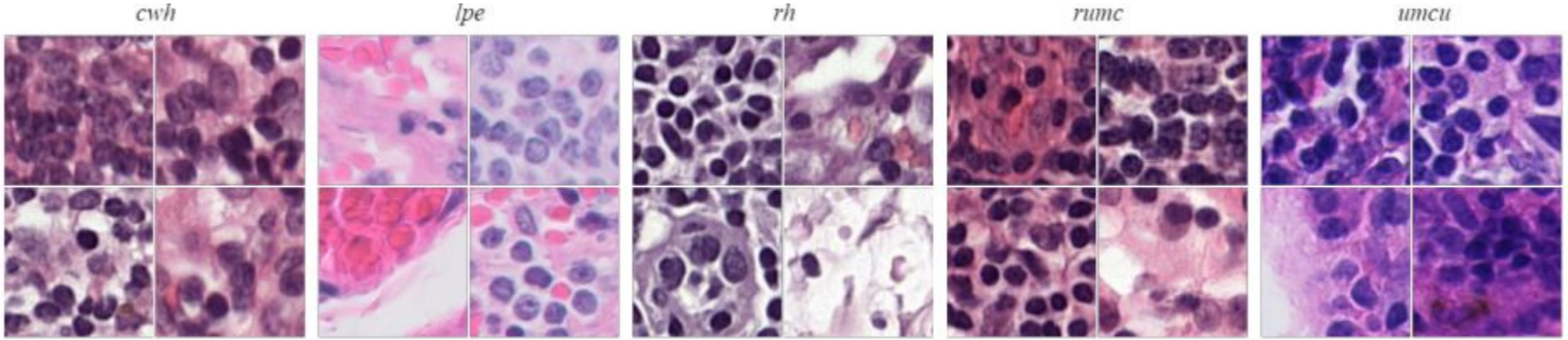
What type of shift is this?

Example: Background

What type of shift is this?

Source: Aniraj, Masking Strategies for Background Bias Removal in Computer Vision Models, 2023

Example: Laboratories



Source: Faryna, Tailoring automated data augmentation to H&E-stained Histopathology, 2021

What type of shift is this?

Example: Crop Yield Forecasting

Different management practices

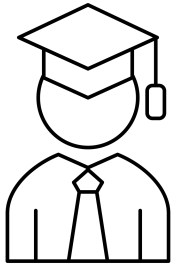
What type of shift is this?

Example: Wildlife Monitoring in Dry vs. Wet Season

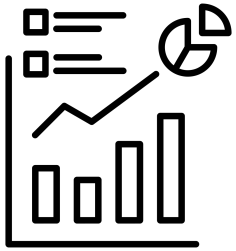
Source: Szenicer, Seismic savanna: machine learning for classifying wildlife and behaviours using ground-based vibration field recordings, 2021

What type of shift is this?

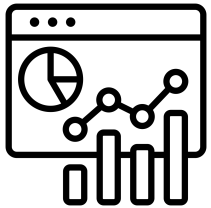
How to Identify Distribution Shift



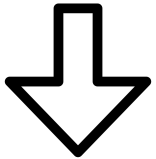
Working with Domain Experts



Visualization Techniques



Statistical Methods



Decreased Inference Performance

How to Identify Distribution Shift: Domain Experts

Understand how data was created and collected

Visual inspection to identify subtle changes

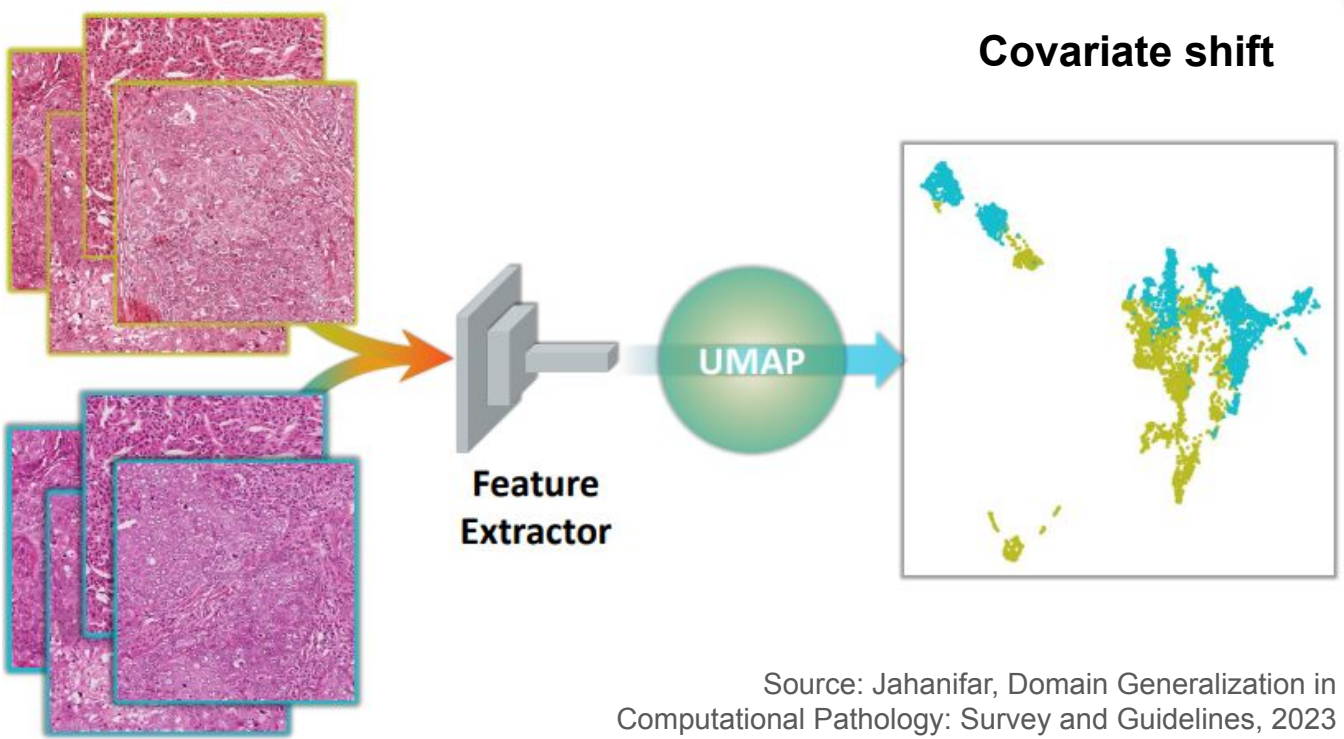
Image quality

Annotation quality

Known variations

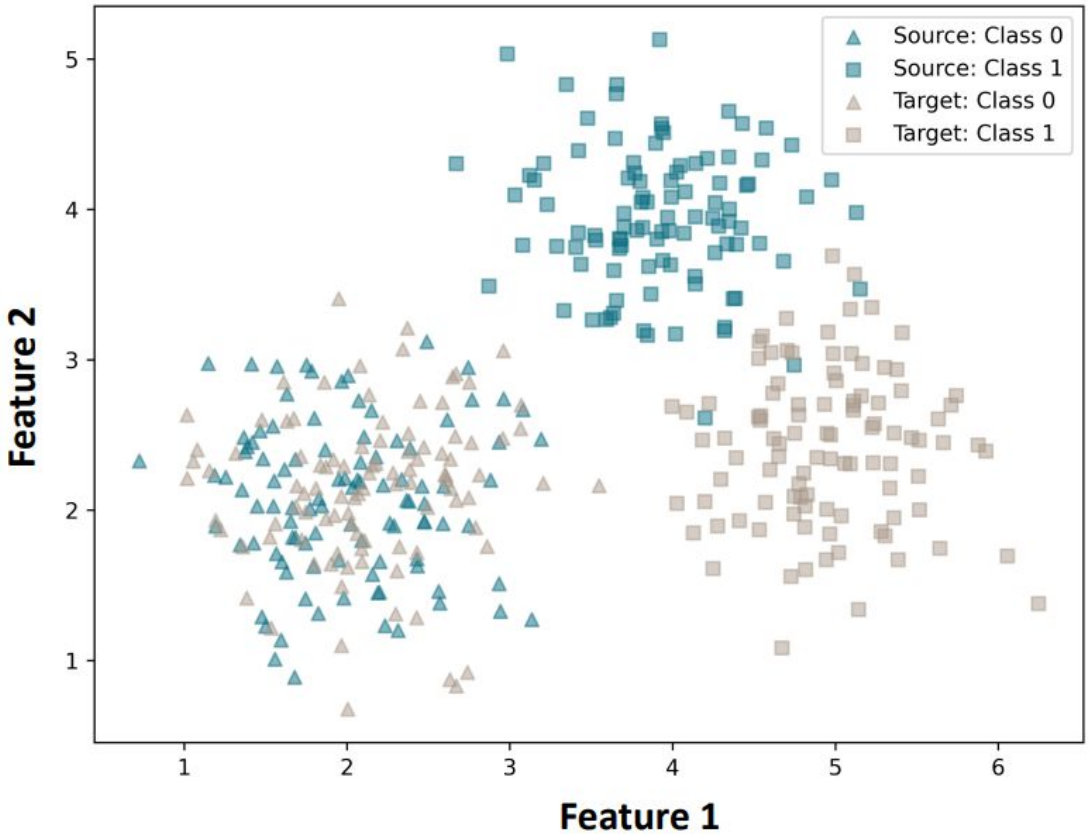
How to Identify Distribution Shift: Visualization Techniques

t-SNE or UMAP for visualizing high-dimensional data distributions



Source: Jahanifar, Domain Generalization in Computational Pathology: Survey and Guidelines, 2023

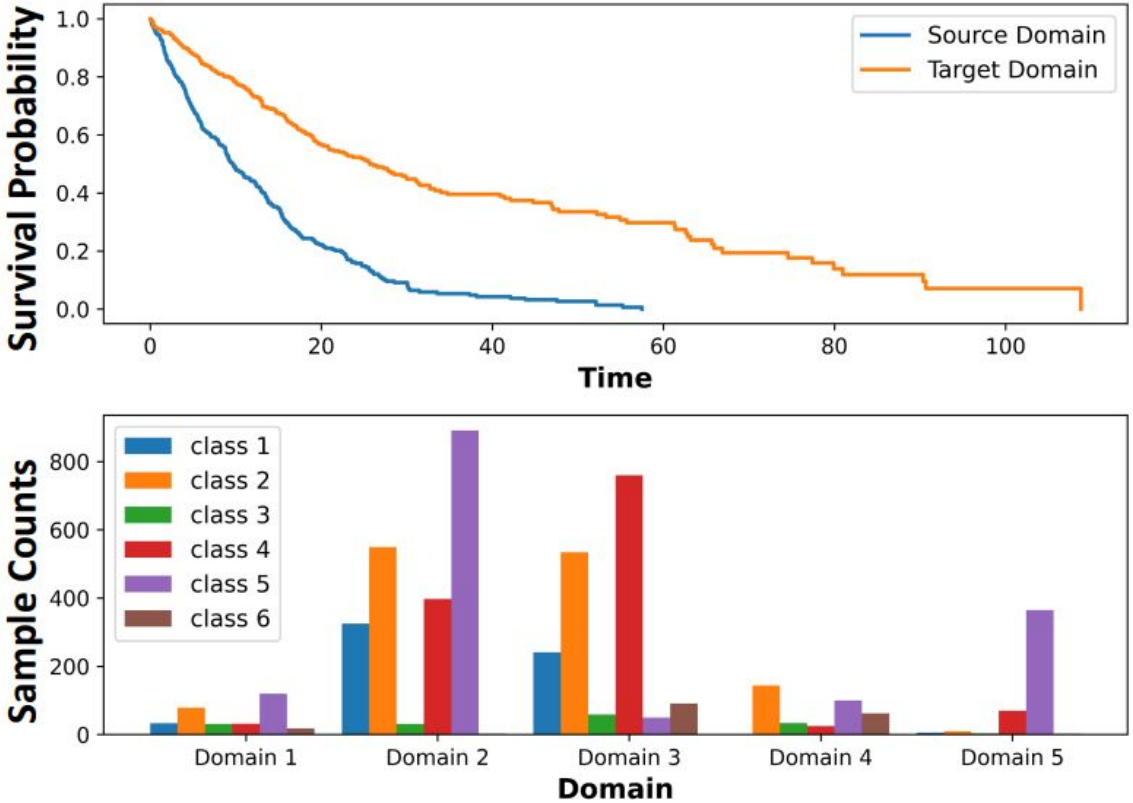
How to Identify Distribution Shift: Visualization Techniques



Class-conditional shift

Source: Jahanifar, Domain Generalization in Computational Pathology: Survey and Guidelines, 2023

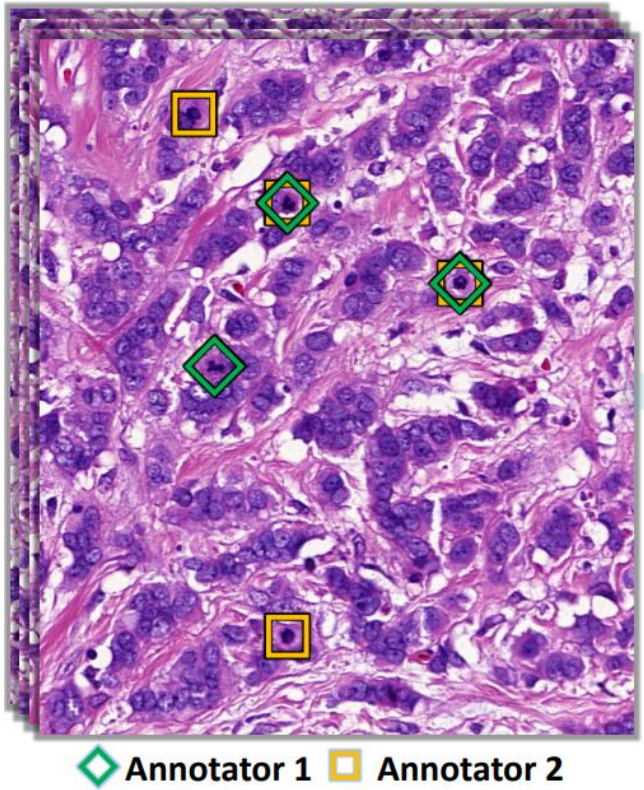
How to Identify Distribution Shift: Statistical Methods

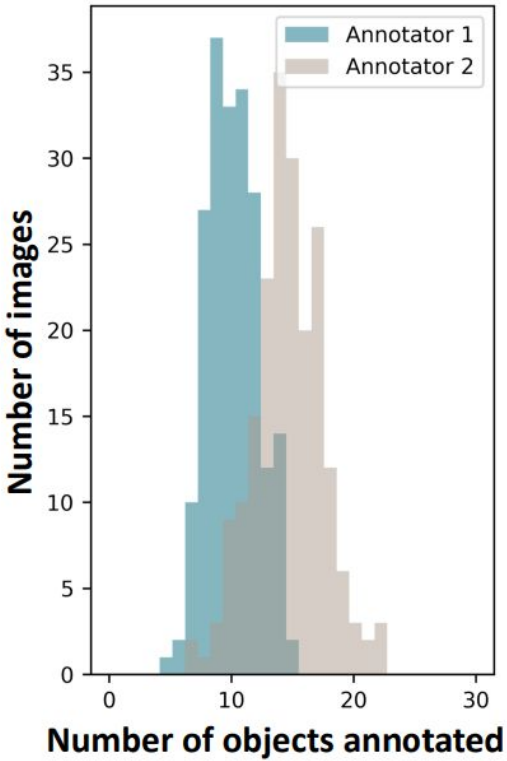


Prior shift

Source: Jahanifar, Domain Generalization in Computational Pathology: Survey and Guidelines, 2023

How to Identify Distribution Shift: Statistical Methods

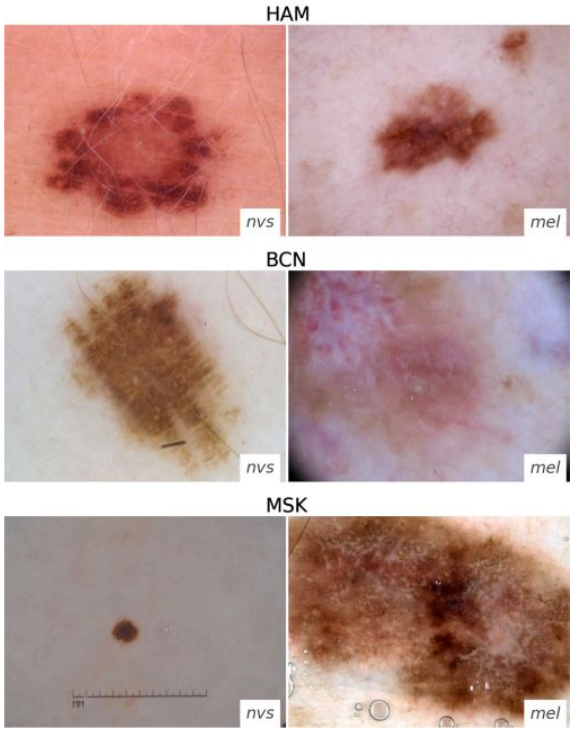




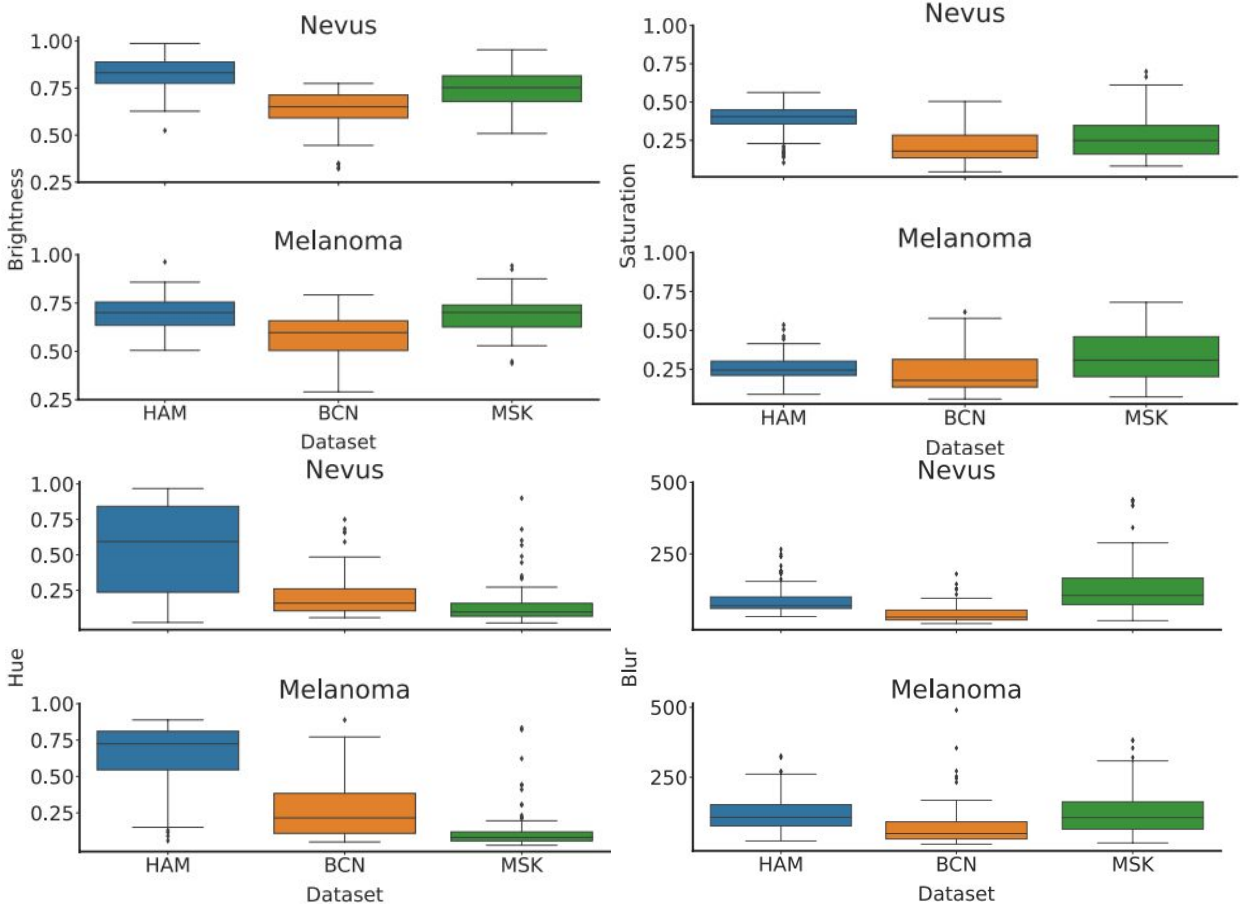
Posterior shift

Source: Jahanifar, Domain Generalization in Computational Pathology: Survey and Guidelines, 2023

How to Identify Distribution Shift: Statistical Methods



Covariate shift

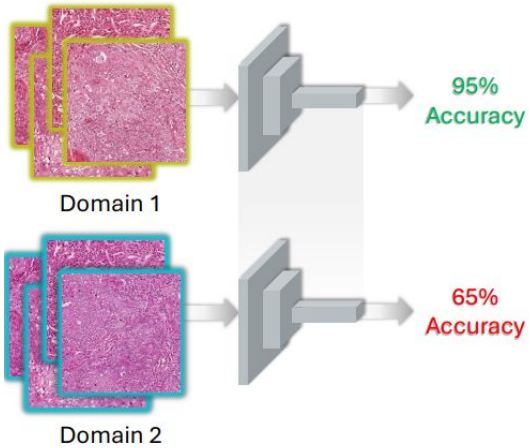


Source: Fogelberg, Domain shifts in dermoscopic skin cancer datasets: Evaluation of essential limitations for clinical translation, 2023

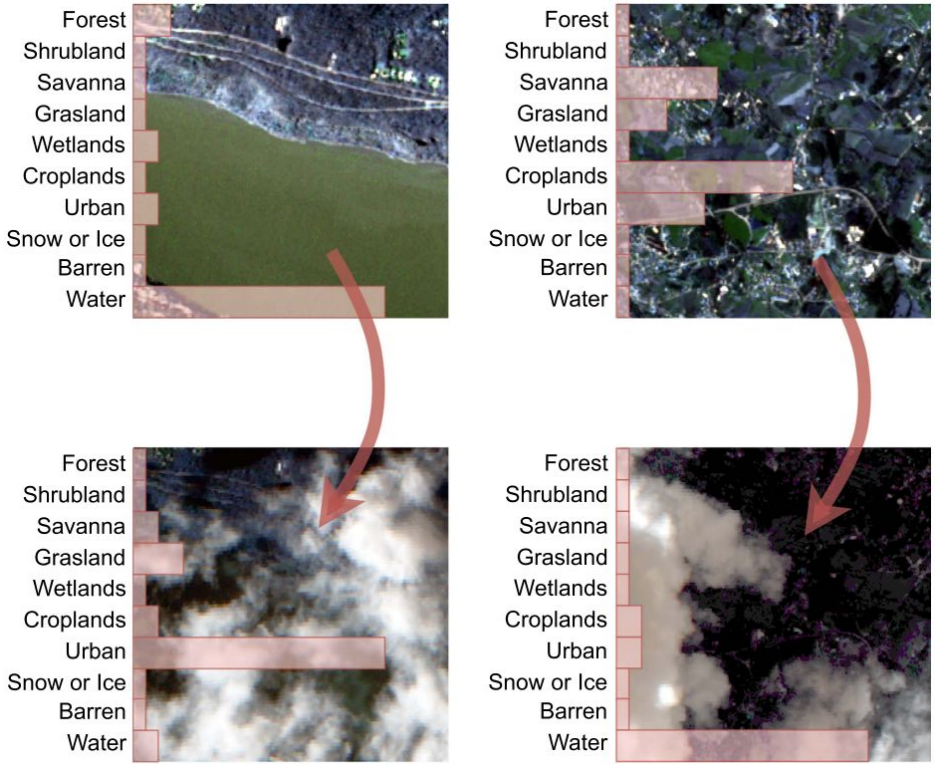
How to Identify Distribution Shift: Decreased Inference Performance

Monitor model performance metrics over time

Compare performance on different subsets of data



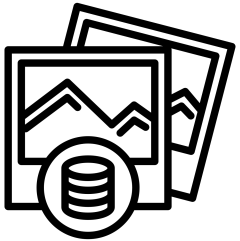
Source: Zamanitajeddin, Benchmarking Domain Generalization Algorithms in Computational Pathology, 2024



Source: Gawlikowski, Explaining the Effects of Clouds on Remote Sensing Scene Classification, 2022

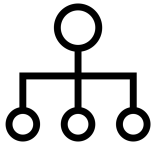
Addressing Distribution Shift

Understand Data Before Modeling



Data-Centric Approaches

Robust Model Development



Structure Test Data

Continuous Monitoring

Addressing Distribution Shift: Understand Data Before Modeling

Identifying potential sources of distribution shift in your domain

- Talk to domain experts
- Understand how data was created and collected

Thorough exploratory data analysis

- Visualization and statistical analysis

Addressing Distribution Shift: Data-Centric Approaches

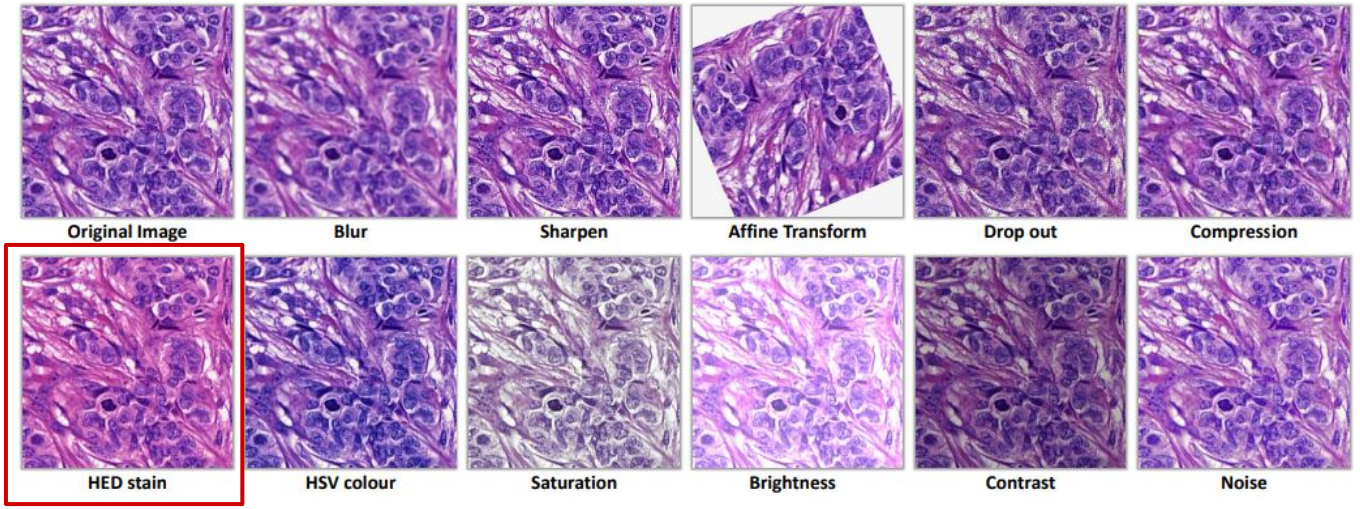
Diverse training set

Image quality control for training and inference

Consistent annotation process

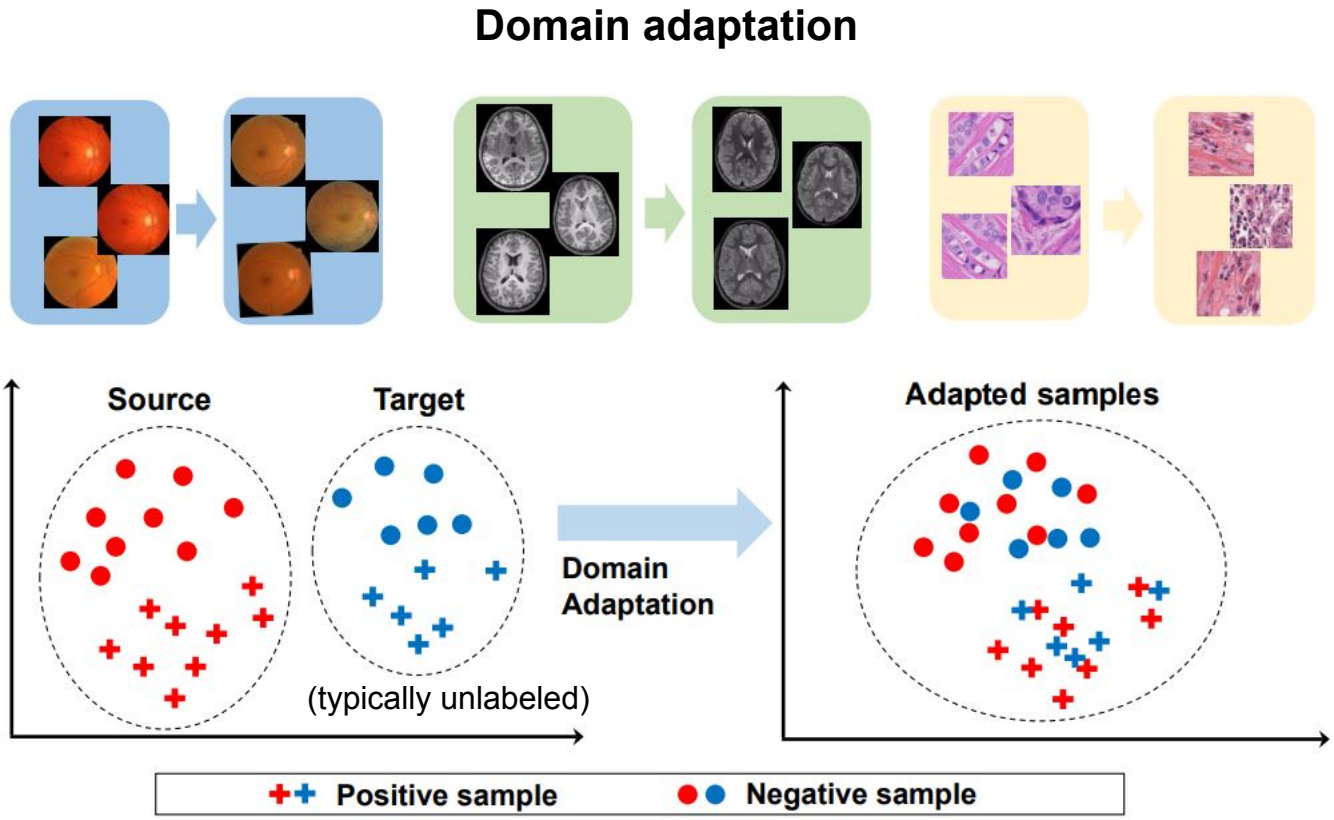
Addressing Distribution Shift: Data-Centric Approaches

Data augmentation to simulate potential shifts



Source: Jahanifar, Domain Generalization in Computational Pathology: Survey and Guidelines, 2023

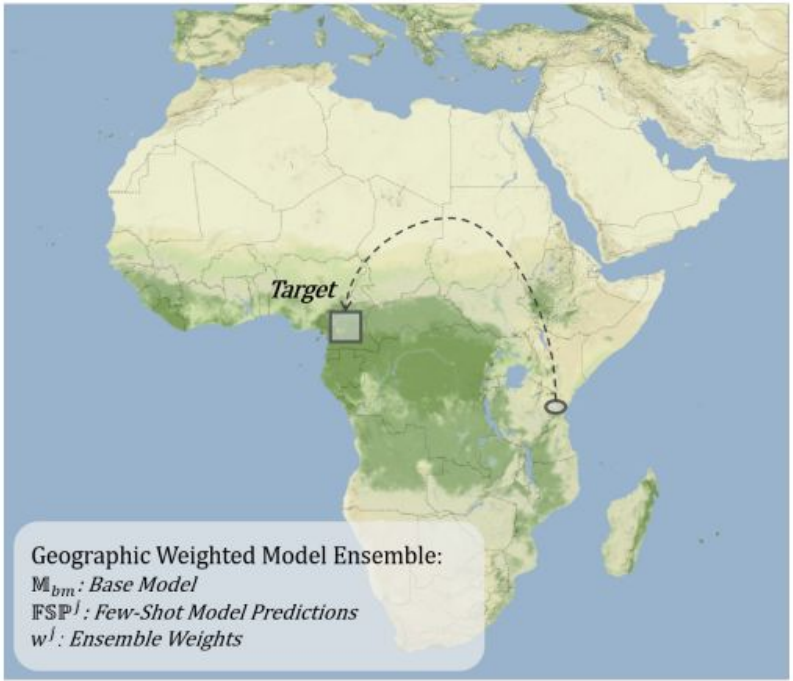
Addressing Distribution Shift: Robust Model Development

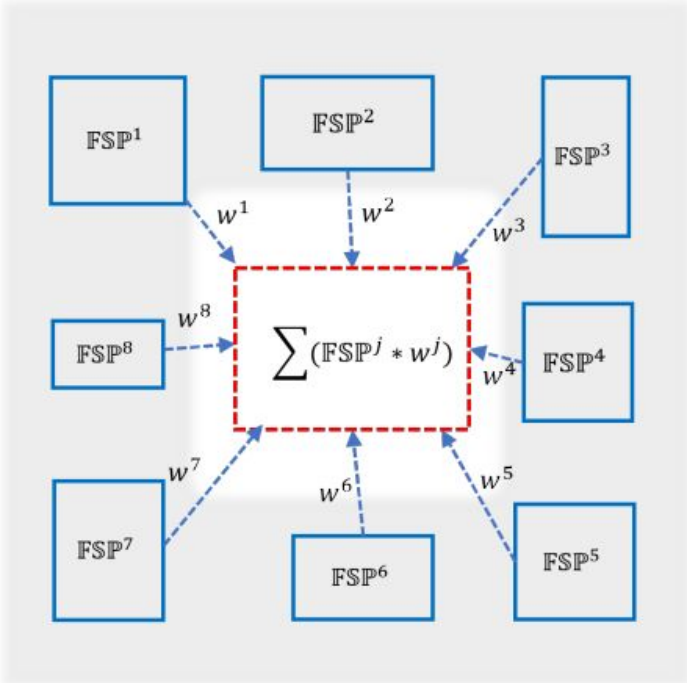


Source: Guan, DomainATM: Domain Adaptation Toolbox for Medical Data Analysis, 2022

Addressing Distribution Shift: Robust Model Development

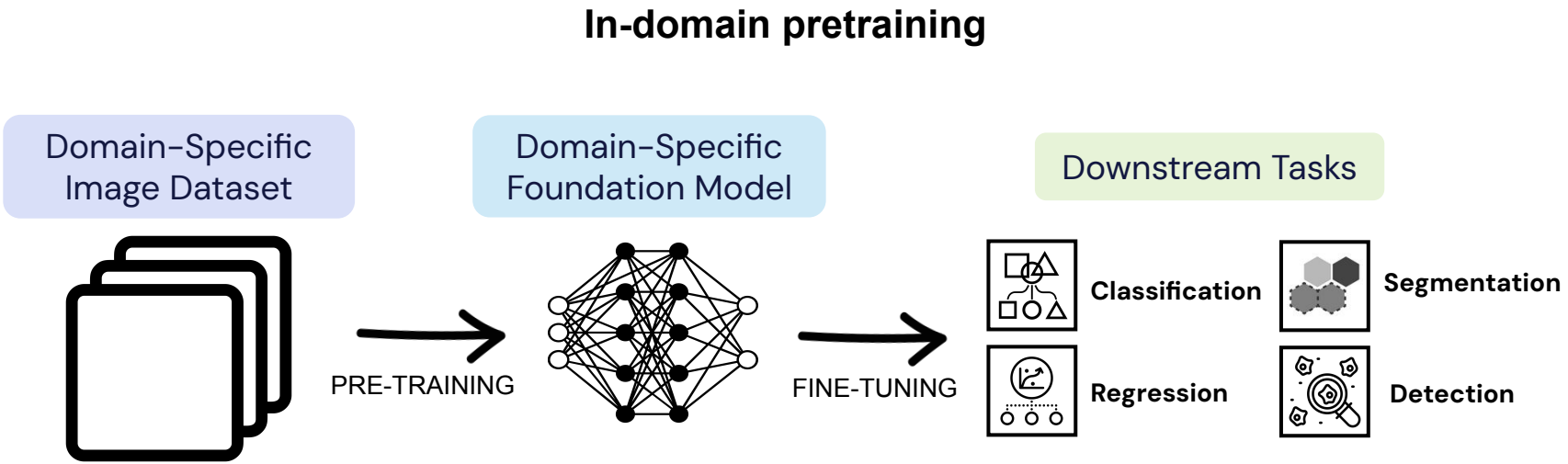
Ensemble methods





Source: Rethink Geographical, 2023 Generalizability with Unsupervised Self-Attention Model Ensemble: A Case Study of OpenStreetMap Missing Building Detection in Africa

Addressing Distribution Shift: Robust Model Development

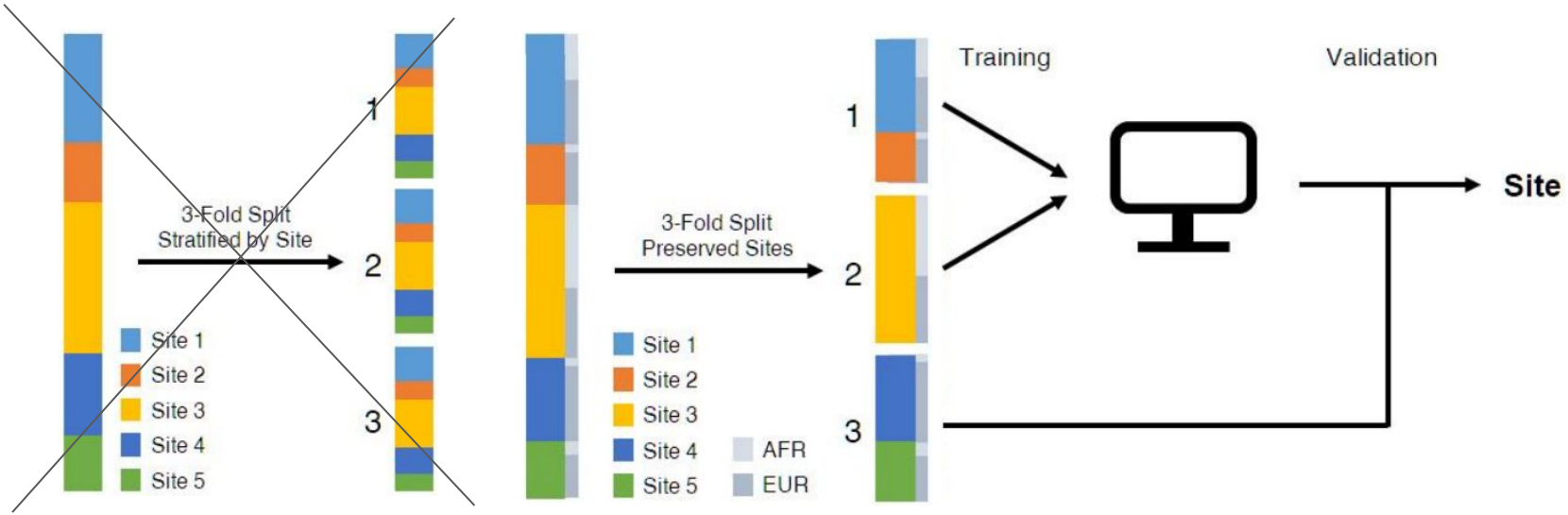


Addressing Distribution Shift: Structure Test Data

Create diverse test sets that represent different scenarios

Regularly update test sets to reflect evolving real-world conditions

Design experiments to be free of data leakage



Source: Howard, The Impact of Digital Histopathology Batch Effect on Deep Learning Model Accuracy and Bias, 2020

Addressing Distribution Shift: Continuous Monitoring

- Performance tracking
- Input and output monitoring
- Anomaly detection
- Model retraining

Summary

- There are multiple types of distribution shift
- Implications: decreased accuracy, bias, unexpected behavior, safety concerns
- Try to identify shift before modeling
- Techniques to address shift depend on type of shift

Computer Vision Consulting

Who I work with:

- Founders and other leaders
- Their technical team

Example results:

- A roadmap to streamline model development
- Break through roadblocks
- Keep up with AI trends and innovations
- Boost investor confidence

Advisory services:

- Monthly strategy call
- Weekly office hours
- Private Slack channel
- And more

Bonus Offer

Team Workshop: Mastering Distribution Shift in Computer Vision

In just 90 minutes, you'll gain:

- Deep insights into the causes and types of distribution shift
- Practical strategies for detecting and mitigating its impact on your data
- A concrete action plan for building models that perform reliably

$1,500 (25% discount if booked in the next 48 hours)

Resources

Team Workshop: Mastering Distribution Shift in Computer Vision

(25% discount if booked in the next 48 hours)

http://pixelscientia.com/workshops/distribution-shift

Other consulting services: heather@pixelscientia.com

Pixel Scientia LABS