© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1719292

Meta-Learning for Domain Generalization Under Distribution Shift: Methods, Benchmarks, And Open Challenges

ARYANIL ROY[1], MAINAK GHATAK[2], KAUSHIK BANERJEE[3], SANANDA CHATTERJEE[4]

[1, 2, 3, 4] *Department of Computer Science and Engineering, Institute of Engineering and Management Kolkata, Kolkata, West Bengal, India*

Abstract- ***Domain generalization (DG) explores how a model trained on a fixed set of source domains can perform reliably on unseen tar-get domains. Meta-learning addresses this by training models to explic-itly learn to generalize, yet curated benchmarks have repeatedly shown that well-tuned evidence-based risk minimization (ERM) is a formidable baseline. This survey examines that tension with a focus on what ac-tually generalizes and why. We contribute three things. First, a bivari-ate taxonomy that cross references shift type (covariate, conditional, in-variant covariate, category, compound) against meta intervention level (data/augmentation, representation/gradient, optimization/parameter, prompt/foundation). Second, a structured comparison of benchmark fam-ilies: DomainBed, WILDS, single source, and open set settings that shows how benchmark choice, and not just method design, drives published conclusions. Third, a critical analysis of conditions under which meta-learning gains over ERM are real versus aphemeral, updated for the 2023-2026 period when foundation models, causal approaches, and meta prompting have substantially changed the landscape. We conclude with actionable open challenges and directions where meta-learning retains a genuine edge.***

Keywords: Domain Generalization, Meta-Learning, Distribution Shift, Domainbed, WILDS, Benchmark Comparison, Transfer Learning

I. INTRODUCTION

A model that generalizes across domains must accommodate shifts in input statistics, label relationships, or both. Standard empirical risk minimization as-sumes training and test data come from the same distribution: an assumption that breaks in the wild. Domain generalization eases it by training on multiple source domains and evaluating on unobserved target domains with no access to target data at training time.

Meta-learning offers a natural fit: if generalization itself can be treated as a skill, then a model can be trained on synthesized train-test splits to develop that skill. The foundational work of Li et al. [14] operationalized this intuition as MLDG, training a model on meta train domains and evaluating update qual-ity on meta test domains in every outer loop.

Since 2018, the landscape has expanded considerably: gradient matching [20], causal meta-learning [3], meta-regularization [2], and, most recently, meta-prompting for frozen foundation models [4].

Two findings complicate this picture. Gulrajani and López-Paz [10] showed that under a controlled evaluation protocol with consistent architectures and model selection, ERM achieves 67.0% on the DomainBed suite while MLDG achieves 66.8%.

Koh et al. [12] introduced the WILDS benchmark, where dis-tribution shifts are real and compound, temporal, geographic, and demographic factors combine, and found standard robust optimization often fails. These two datasets tell almost opposite stories.

This survey is organized around that gap. We do not position it as the first survey on meta-learning for DG; comprehensive reviews by Gholamzadeh Khoee et al. [9] and Hospedales et al. [11] provide thorough coverage through 2024.

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1719292

Instead, we emphasize a shift-aware taxonomy, a structured benchmark compar-ison, a critical treatment of when meta-learning gains are genuine, and develop-ments from 2023–2026 including compound real-world shifts and the transition from weight-optimization to foundation-model adaptation.

II. BACKGROUND AND SCOPE

2.1 Problem Formulation

Let D = {D1, . . ., DK} be a collection of source domains, each with distribution Pk (X, Y). The goal of domain generalization is to learn a predictor fθ from D that minimizes expected risk on an unseen target domain Dtest whose distribu-tion P test is drawn from the same meta-distribution but not observed during training. The key difficulty is that Pk differ from each other and from P test in ways that are only partially known.

Distribution shift takes several distinct forms. Covariate shift occurs when P (X) changes but P (Y |X) remains stable. Conditional shift occurs when P (Y |X) itself changes. Invariant covariate shift [26] occurs when even the marginal distri-bution of causally invariant features moves across domains. Category shift arises when source and target label spaces differ [30]. Compound shift, the kind preva-lent in WILDS [12], combines temporal, geographic, and demographic factors simultaneously.

2.2 Meta-Learning Formulation for DG

In the MLDG formulation [14], training proceeds iteratively. At each episode, source domains are split into meta train domains Dtr and meta test domains Dte. A fast adaptation step computes $\theta' = \theta - \alpha \nabla_\theta L_{Dtr}(\theta)$. The meta objective then evaluates the refined parameters on Dte:

$$\min_\theta L_{Dtr}(\theta) + \beta L_{Dte}(\theta') \quad (1)$$

This simulates the train test domain gap during training, inducing the model to learn features that remain useful after a domain shift. Gradient matching methods such as Fish [20] instead seek parameters where per domain gradients are aligned, emulating the same objective more efficiently.

III. SHIFT-AWARE TAXONOMY

We organize meta-learning methods for DG along two axes. Figure 1 illustrates the structure.

3.1 Axis 1: Shift Type Addressed

Covariate/diversity shift. Style, microstructure, or sensor level changes where class conditional distributions remain stable. Most DomainBed datasets (PACS, VLCS, OfficeHome) fall here. Methods such as MLDG [14], Fish [20], and SAM-GM [23] target this scenario.

Conditional/correlation shift. Incidental correlations between labels and context features differ across domains. DecAug [1] uses gradient orthogonaliza-tion to isolate category features from incidental context. Adversarial IRM [29] unifies invariant risk optimization with domain wise adversarial training.

Invariant covariate shift. Wong et al. [26] identify a failure mode missed by conventional gradient alignment: even invariant features can have different marginals across domains, causing methods that ignore feature density to de-grade. Weighted risk invariance directly remedies this.

Category/open-set shift. Source and target label spaces differ. DAML [30] uses Dirichlet mixup with distilled soft labels for open-domain DG. Dualistic meta-learning [24] matches gradients across both domains and classes to regu-larize boundaries against unknown-class rejection errors.

Compound/real-world shift. Temporal, geographic, and demographic fac-tors shift simultaneously. WILDS [12] curate's datasets like iWildCam and Came-lyon17 with these properties. Visual domain prompt generation [4] targets this setting for frozen foundation models.

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1719292

3.2 Axis 2: Meta-Intervention Level

Data/augmentation level. Meta-learning learns which augmentations or do-main perturbations to apply. Methods include adversarial task augmentation [22], pseudo multi-source generation from a single domain [7], DecAug [1], AdvST [31], and meta-learned augmentation for histopathology [8].

Representation/gradient level. Gradient alignment across domains dur-ing the meta-objective. Fish [20] provides an efficient first-order approximation. SAM-GM [23] adds sharpness-aware minimization. Dualistic meta-learning [24] extends matching across both domains and classes. Causality-inspired mask-ing [17] enforces causal sufficiency in representations.

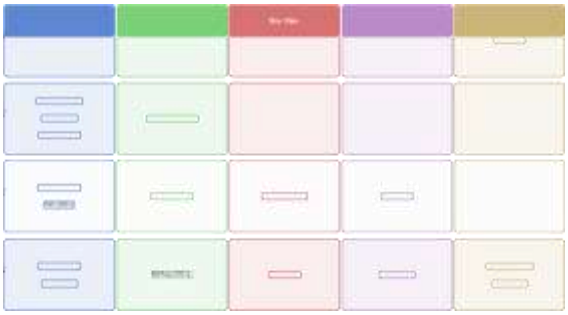

Fig. 1. Two-axis taxonomy of meta-learning approaches for domain generalization. Shift type (horizontal) crosses meta-intervention level (vertical). Methods are placed by their primary mechanism. Foundation-model methods occupy the rightmost shift column since they are designed for compound real-world scenarios.

Optimization/parameter level. Episodic optimization of initialization or specific parameters. MLDG [14] and MetaReg [2] operate at this level. Per-sample adaptation via amortized variational inference [28] and MetaVIB [6] treat classifier parameters probabilistically. MABN [27] meta-learns only batch nor-malization affine parameters, preventing knowledge interference during test-time adaptation.
Prompt/foundation level. With frozen large-scale backbones, the meta-intervention is a learned prompt generator. Visual domain prompt generation (VPG) [4] maintains a knowledge bank and produces domain-specific prompts for CLIP-style models without weight updates, a qualitative shift from the MLDG-era paradigm.

IV. METHODS AND BENCHMARK LANDSCAPE

4.1Representative Methods by Family
Table 1 summarizes key methods across families with their benchmark settings and primary shift targets.

Gradient matching. Fish [20] rewrites the inter-domain gradient alignment
Σ
objective as a tractable first-order update: $\max\theta$

$$\langle \nabla_\theta L^i, \nabla_\theta L^j \rangle, \quad \text{approxi-}$$

$i=j$
mated without second-order terms. SAM-GM [23] adds sharpness-aware per-turbations to ensure alignment targets flat minima rather than sharp ones,

Table 1. Representative meta-learning methods for domain generalization. Setting: MDG = multi-source, SDG = single-source, ODG = open-domain, OSDG = open-set. Shift: Cov = covariate, Cond = conditional, Mix = compound.

Method	Year	Family	Setting	Shift
MLDG [14]	2018	Opt/episodic	MDG	Cov
MetaReg [2]	2018	Regularizer	MDG	Cov
MetaVIB [6]	2020	Probabilistic	MDG	Cov
Fish [20]	2021	Grad match	MDG	Mix
DecAug [1]	2021	Grad+Aug	MDG	Cond
DAML [30]	2021	Aug/ODG	ODG	Cat
Meta-causal [3]	2023	Causal	SDG	Cov
SAM-GM [23]	2023	Grad match	MDG	Cov
MABN [27]	2024	Arch. decoup.	MDG	Cov
WRI [26]	2024	Opt/weighted	MDG	Mix
VPG [4]	2024	Meta-prompt	MDG	Mix
DML [24]	2023	Grad/OSDG	OSDG	Cat
MATS [19]	2023	Task sampling	MDG	Cov
PseudoDG [7]	2025	Aug/SDG	SDG	Cov

which matters on DomainNet and TerraIncognita. Dong et al. [5] provide an

information-theoretic proof that gradient alignment and representation align-ment are complementary, not competing, strategies—a result that resolves years of empirical ambiguity about which to prefer.

Causal meta-learning. Meta-causal learning [3] replaces the gradient-alignment objective with a simulate-analyze-reduce paradigm: a causal graph generates counterfactual samples, the model analyzes which features exhibit causal stabil-ity, and adaptation weights are assigned accordingly. Causality inspired repre-sentation learning [17] uses adversarial masking to enforce causal sufficiency as a meta regularizer. These approaches are fundamentally more principled than gra-dient alignment for conditional shift, but their computational overhead is higher and their superiority on standard benchmarks remains modest.

Meta-prompting and foundation models. VPG [4] represents a quali-tative change in the meta-learning paradigm. Rather than optimizing weight initializations, a lightweight prompt generator is meta-trained while the back-bone (CLIP or similar) stays frozen. The generator produces domain-specific visual prompts from a learnable knowledge bank. On WILDS, this approach matches or exceeds full fine-tuning with substantially fewer trainable parame-ters. AdvST [31] frames standard data augmentations as semantics transforma-tions learned through adversarial distributionally robust optimization, closing the gap between augmentation-based and optimization-based approaches.

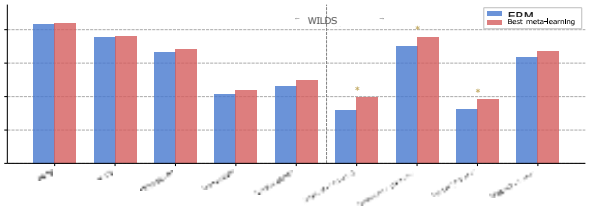

Fig. 2. Schematic benchmark landscape for domain generalization. Positions summa-rize how benchmark families are typically used in the literature rather than exact measured scores. Classical DomainBed benchmarks cluster in the lower-left region, while WILDS and selected scientific benchmarks occupy the higher-realism and higher-heterogeneity region.

Single-source and federated settings. When only one source domain is avail-able, multi-source methods cannot be applied directly. PseudoDG [7] generates synthetic pseudo-domains from a single source, enabling DomainBed-style multi-source algorithms.

Notably, gradient-based meta-learning can deteriorate when meta-train/test splits rely on style-transferred pseudo-domains rather than gen-uinely distinct domains. In federated settings, Multi-Source CGDM [25] adapts gradient discrepancy minimization for decentralized clients without sharing raw data, extending gradient matching to privacy-constrained scenarios.

Non-vision extensions. Meta-learning for DG extends beyond images. GNN initialization via MAML on molecular property prediction [18] handles chemi-cal task heterogeneity. MLDGG [21] integrates causal disentanglement for graph topological shift. Industrial time-series fault diagnosis [16] uses episodic train-ing with feature orthogonality. Medical imaging work [13] addresses task level overfitting when source domains are scarce clinical centres.

4.2 Benchmark Landscape

Figure 2 summarizes the benchmark landscape used in this survey. The hori-zontal axis is a schematic notion of shift realism: benchmarks on the left mainly encode stylized appearance change, whereas those on the right involve geograph-ically, temporally or institutionally entangled shifts. The vertical axis reflects the domain heterogeneity which is the difference in sensor, collection protocol, label regime or structural modality.

On the standard DomainBed benchmarks, ERM with a fixed backbone and controlled model selection closes most of the gap to dedicated meta-learning methods – this is the emerging pattern across the literature. According to Gul-rajani and López-Paz [10], the average MLDG across the DomainBed setting was at 66.8%, whereas the average ERM result was at 67.0%. The patterned covariate shifts in

these benchmarks are still informative, but they are weak proxies for the compound shifts found in hospital, ecological and remote-sensing deployments.

WILDS [12] alter the evaluation narrative. Using iWildCam, Camelyon17 and FMoW datasets, gains reported from distribution-shift-aware methods are more pronounced, especially for gradient-aligning methods across domains [20] or frozen pretrained backbone-adapting via prompt generation [4]. The specific margin depends on the metric and validation protocol, but the high-level take-away is robust: the advantage of meta-learning increases when the target shift is genuine compound rather than just stylistic.

V. CRITICAL ANALYSIS AND OPEN CHALLENGES

5.1 When Meta-Learning Beats ERM—and When It Does Not

The DomainBed result [10] is reproducible and should be taken seriously. Under matched architectures and model selection, the episodic training signal of MLDG adds little to what a well-regularized ERM already achieves on stylized covariate shifts.

The reasons are not mysterious. Standard augmentation, weight decay, and batch normalization act as implicit domain-invariance regularizers. Episodic training provides an explicit signal, but the domain partition it simulates is artificial—meta-train and meta-test domains are sampled from the same small set of source domains.

The picture changes in three identifiable conditions:
Large domain count. On DomainNet (six domains, 600K images), gradient matching methods like SAM-GM [23] show consistent 1–3% gains over ERM. With more domains, simulated meta-test splits are less redundant.

Compound real-world shifts. On WILDS datasets, where shifts are temporal, geographic, and demographic simultaneously, explicit adaptation mechanisms—including meta-learned initialization, test-time batch normalization [27], and meta-prompting [4]—provide measurable gains over ERM.
Architectural targeting.

Meta-learning only specific model components, such as batch normalization affine parameters (MABN [27]), can prevent the knowl-edge interference that afflicts full-network episodic training. This targeted ap-proach works well in test-time adaptation settings where a small amount of unlabeled target data is available.

Conversely, in single-source DG, gradient-based meta-learning can deterio-rate when the meta-train/test split is constructed from style-transferred pseudo-domains [7]. The simulated shift is synthetic and may not reflect the structure of real test-time shifts, so the episodic signal trains the model on misleading variation.

5.2 Benchmark Choice Changes Conclusions

The discrepancy between DomainBed and WILDS conclusions is not incidental—it reflects a fundamental difference in what is being tested. PACS-style bench-marks measure robustness to artistic style, a factor that proper augmentation can largely address. WILDS benchmarks measure robustness to shifts in deploy-ment context that cannot be anticipated through augmentation alone.

This has a practical implication for reading the literature: a method reported to outperform ERM by 3% on PACS may not translate to any gain on iWildCam. The converse is also true, methods tuned for real world compound shifts may not improve on stylized benchmarks where ERM already performs well.

The 2025 PseudoDG result [7] adds a further wrinkle: generating synthetic pseudo-domains from a single source to enable multi-source DG algorithms does not reliably help when the base algorithm relies on gradient-based meta-learning. The style-transfer process introduces an artificial domain signal that the meta-learner overfits to, a form of task-level overfitting [13].

5.3 The Foundation-Model Transition

Since 2022, the weight-optimization paradigm underlying MLDG and Fish has been complemented by a prompt-optimization paradigm. Rather than learning domain-agnostic initializations, VPG [4] and

related methods learn to generate domain-specific prompts for frozen CLIP-style backbones. This is not a marginal extension; it is a different approach to the same problem.

The advantage of meta-prompting is that the backbone's learned representations are preserved. Standard fine-tuning on source domains can degrade CLIP's zero-shot generalization to novel target domains. Meta-learning a prompt generator avoids this by keeping backbone weights fixed.

The remaining open ques-tion is whether the prompt generator itself overfits to the distribution of source domains it was trained on—a question analogous to the task-level overfitting problem in episodic training.

5.4 Open Challenges

Challenge 1: ERM as a persistent baseline. Meta-learning methods rarely outperform ERM by more than 2–3% on stylized benchmarks under controlled evaluation [10]. Identifying conditions under which episodic training provides reliable gains—and designing methods targeted to those conditions—remains the central theoretical problem.

Challenge 2: Invariant covariate shift. Wong et al. [26] show that stan-dard gradient alignment fails when the marginal distribution of causally invari-ant features differs across domains. Existing methods largely ignore feature-level density mismatch. Extending meta-learning objectives to account for invariant covariate shift is an open problem with real-world relevance.

Challenge 3: Task-level overfitting in episodic training. When source domains are few, simulated meta-test splits carry little information beyond the meta-train splits. Methods like mixed task sampling [13] and adaptive task sam-pling [19] partially address this, but a principled solution for the low-source-diversity regime remains elusive.

Challenge 4: Prompt generator generalization. Meta-prompting avoids backbone overfitting but shifts the risk to the prompt generator itself. Whether prompt generators trained on a small set of source domains can produce useful prompts for substantially different target domains is not yet well understood.

Challenge 5: Computational overhead of bi-level optimization. Com-puting Hessian-vector products can be expensive at scale, which is the compu-tational overhead of bi-level optimization. Fish [20] sidesteps this with a first-order approximation, however, when the domains are many or heterogeneous this approximation is less faithful. The compound shift benchmarks necessitate an effective meta-learning framework for large scale adoption.

Challenge 6: Open-set and open-domain generalization. When the target domain contains classes not present in the source domains, one vs all classifiers have boundary bias that misclassifies known classes in unknown do-mains [24]. This issue is somewhat addressed by using dualistic gradient match-ing. However, open set benchmarks have a huge gap with well supervised closed set performance.

Challenge 7: Privacy-constrained federated DG. The process of adapt-ing gradient-matching objectives to decentralized settings [25] involves sharing gradient statistics that might contain privacy-sensitive information. Federated DG with strong privacy guarantees is an open problem; existing approaches work under limited threat models.

Challenge 8: Dynamic and test-time shifts. Many meta-learning strate-gies presume there is a constant testing pattern of distribution. Actual deploy-ments undergo constant shifts. Adaptation at the time of testing [15] as shown makes use of unlabeled batches present at the target, however that raises the questions of latency and privacy. One of the rising directions that is underex-plored is meta-initialization with streaming test time adaptation.

CONCLUSION

Meta-learning for domain generalization has progressed considerably since MLDG, but its relationship to strong ERM baselines remains a nuanced one. On curated benchmarks with few source domains, the gains from episodic training are often marginal. On benchmarks with heterogeneous real-world shifts and large domain counts, meta-learning provides genuine and reproducible

improvements.

The developed shift-aware taxonomy in this survey provides a useful lens to evaluate claims: a method's performance on covariate-shift benchmarks re-veals little about its behavior under compound shifts; and the choice of bench-mark should be reported with method design when generalization conclusions are drawn. The shift to foundation models has not removed the necessity for meta-learning; rather it has changed the point of intervention from weight ini-tialization to prompt generation, which preserves the core intuition of learning to adapt and not destructively fine-tune.

Where meta-learning still has the clearest edge: compound real-world shifts with large domain count (WILDS-style), open-set generalization, and foundation-model adaptation via prompting. Where it does not clearly win: stylized covariate-shift benchmarks where ERM with proper regularization is already near the ceiling. Acknowledging this distinction is, arguably, the most practically useful contribution a survey on this topic can make.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

REFERENCES

[1] Bai, H., Sun, R., Hong, L., Zhou, F., Ye, N., Ye, H.J., Chan, S.H.G., Li, Z.: DecAug: Out-of-distribution generalization via decomposed feature representation and se-mantic augmentation. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence (2021). https://doi.org/10.1609/aaai.v35i8.16829

[2] Balaji, Y., Sankaranarayanan, S., Chellappa, R.: MetaReg: Towards domain gener-alization using meta-regularization. In: Advances in Neural Information Processing Systems. vol. 31 (2018)

[3] Chen, J., Gao, Z., Wu, X., Luo, J.: Meta-causal learning for sin-gle domain generalization. arXiv preprint arXiv:2304.03709 (2023). https://doi.org/10.48550/arXiv.2304.03709

[4] Chi, Z., Gu, L., Zhong, T., Liu, H., Yu, Y., Plataniotis, K.N., Wang, Y.: Adapt-ing to distribution shift by visual domain prompt generation. arXiv preprint arXiv:2405.02797 (2024). https://doi.org/10.48550/arXiv.2405.02797

[5] Dong, Y., Gong, T., Chen, H., Song, S., Zhang, W., Li, C.: How does distribu-tion matching help domain generalization: An information-theoretic analysis. arXiv preprint arXiv:2406.09745 (2024). https://doi.org/10.48550/arXiv.2406.09745

[6] Du, Y., Xu, J., Xiong, H., Qiu, Q., Zhen, X., Snoek, C.G.M., Shao, L.: Learning to learn with variational information bottleneck for domain generalization. arXiv preprint arXiv:2007.07645 (2020). https://doi.org/10.48550/arXiv.2007.07645

[7] Enomoto, S.: Pseudo multi-source domain generalization: Bridging the gap between single and multi-source domain generalization. arXiv preprint arXiv:2505.23173 (2025). https://doi.org/10.48550/arXiv.2505.23173

[8] Faryna, K., van der Laak, J., Litjens, G.: Automatic data augmen-tation to improve generalization of deep learning in h&e-stained histopathology. Computers in Biology and Medicine 170, 108018 (2024). https://doi.org/10.1016/j.compbiomed.2024.108018

[9] Gholamzadeh Khoee, A., Yu, Y., Feldt, R.: Domain generalization through meta-learning: A survey. Artificial Intelligence Review (2024). https://doi.org/10.1007/s10462-024-10922-z

[10] Gulrajani, I., López-Paz, D.: In search of lost domain generalization. arXiv preprint arXiv:2007.01434 (2020). https://doi.org/10.48550/arXiv.2007.01434

[11] Hospedales, T.M., Antoniou, A., Micaelli, P., Storkey, A.: Meta-learning in neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelli-gence 44(9), 5149–5169 (2021). https://doi.org/10.1109/tpami.2021.3079209

[12] Koh, P.W., Sagawa, S., Marklund, H., Xie, S.M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R.L., Gao, I., Lee, T., David, E., Stavness, I., Guo, W., Earnshaw, B., Haque, I., Beery, S.M., Leskovec, J., Kundaje, A., Pierson, E., Levine, S., Finn, C., Liang, P.: WILDS: A bench-mark of in-the-wild distribution shifts. arXiv preprint arXiv:2012.07421 (2020). https://doi.org/10.48550/arXiv.2012.07421

[13] Li, C., Liu, Y., Li, M., Li, X., Song, Y., Yu, Y.: Domain generalization on med-ical imaging classification using episodic training with task augmentation. arXiv preprint arXiv:2106.06908 (2021). https://doi.org/10.48550/arXiv.2106.06908

[14] Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Learning to generalize: Meta-learning for domain generalization. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence (2018). https://doi.org/10.1609/aaai.v32i1.11596

[15] Liang, J., He, R., Tan, T.: A comprehensive survey on test-time adap-tation under distribution shifts. arXiv preprint arXiv:2303.15361 (2023). https://doi.org/10.48550/arXiv.2303.15361

[16] Liao, Y., et al.: Episodic training and feature orthogonality-driven domain gener-alization for rotating machinery fault diagnosis under unseen working conditions. Machines 13(7), 563 (2025). https://doi.org/10.3390/machines13070563

[17] Lv, F., Liang, J., Li, S., Zang, B., Liu, C.H., Wang, Z., Liu, D.: Causality inspired representation learning for domain generalization. arXiv preprint arXiv:2203.14237 (2022). https://doi.org/10.48550/arXiv.2203.14237

[18] Nguyen, C.Q., Kreatsoulas, C., Branson, K.M.: Meta-learning GNN initializations for low-resource molecular property prediction. arXiv preprint arXiv:2003.05996 (2020). https://doi.org/10.48550/arXiv.2003.05996

[19] Shen, Z., Yu, H., Cui, P., Liu, J., Zhang, X., Zhou, L., Liu, F.: Meta adaptive task sampling for few-domain generalization. arXiv preprint arXiv:2305.15644 (2023). https://doi.org/10.48550/arXiv.2305.15644

[20] Shi, Y., Seely, J., Torr, P.H.S., Siddharth, N., Hannun, A., Usunier, N., Synnaeve, G.: Gradient matching for domain generalization. arXiv preprint arXiv:2104.09937 (2021). https://doi.org/10.48550/arXiv.2104.09937

[21] Tian, Q., Zhao, C., Shao, M., Wang, W., Lin, Y., Li, D.: MLDGG: Meta-learning for domain generalization on graphs. arXiv preprint arXiv:2411.12913 (2024). https://doi.org/10.48550/arXiv.2411.12913

[22] Wang, H., Deng, Z.: Cross-domain few-shot classification via adversarial task aug-mentation. In: Proceedings of the 30th International Joint Conference on Artificial Intelligence. pp. 1–7 (2021). https://doi.org/10.24963/ijcai.2021/149

[23] Wang, P., Zhang, Z., Lei, Z., Zhang, L.: Sharpness-aware gradient matching for domain generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023). https://doi.org/10.1109/cvpr52729.2023.00367

[24] Wang, X., Peng, D., Hu, P., Sang, J.: Generalizable decision boundaries: Dualistic meta-learning for open set domain generalization. arXiv preprint arXiv:2308.09391 (2023). https://doi.org/10.48550/arXiv.2308.09391

[25] Wei, Y., Han, Y.: Multi-source collaborative gradient discrepancy minimization for federated domain generalization. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence (2024). https://doi.org/10.1609/aaai.v38i14.29510

[26] Wong, G., et al.: Weighted risk invariance: Domain generalization un-der invariant feature shift. arXiv preprint arXiv:2407.18428 (2024). https://doi.org/10.48550/arXiv.2407.18428

[27] Wu, Y., Chen, Z., Wang, W., Liu, Y.: Test-time

domain adaptation by learning domain-aware batch normalization. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence (2024). https://doi.org/10.1609/aaai.v38i14.29527

[28] Xiao, Z., Shen, X., Zhen, X., van den Hengel, A., Shao, L., Snoek, C.G.M.: Learning to generalize across domains on single test samples. arXiv preprint arXiv:2202.08045 (2022). https://doi.org/10.48550/arXiv.2202.08045

[29] Xin, S., Wang, Y., Li, J., Guo, Y., Ding, P., Li, W.: On the connection between invariant learning and adversarial training for out-of-distribution generalization. In: Proceedings of the 37th AAAI Conference on Artificial Intelligence (2023). https://doi.org/10.1609/aaai.v37i9.26250

[30] Yang, S., Wang, Y., Joao Ribeiro, D., Bhatt, R., Lim, S.N., Torr, P.H.S.: Open domain generalization with domain-augmented meta-learning. arXiv preprint arXiv:2104.03620 (2021). https://doi.org/10.48550/arXiv.2104.03620

[31] Zheng, G., Huai, M., Zhang, A.: AdvST: Revisiting data augmentations for single domain generalization. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence (2024). https://doi.org/10.1609/aaai.v38i19.30184