Archives available at journals.mriindia.com

International Journal on Advanced Computer Theory and Engineering

ISSN: 2319 - 2526

Volume 15 Issue 01s, 2026

Learning To Learn: A Survey of Recent Literature on Meta Learning

[1]Mohan Pramod Patil , [2]Dr. G .K. Patnaik

[1]SSBT's College of Engineering and Technology, Jalgaon, Maharashtra, India 425001.

[2] Principal, SSBT's College of Engineering and Technology, Jalgaon , Maharashtra, India 425001

Email: [1] mhnptl90@gmail.com, [2] Patnaik.girish@gmail.com

Peer Review Information

Submission: 08 Dec 2025

Revision: 25 Dec 2025

Acceptance: 10 Jan 2026

Keywords

Meta Learning, Meta training, Evaluation metrics, AutoML.

Abstract

The essence of Meta-learning is "learning to learn". Meta Learning is a subset of machine learning. Meta-learning is the process of using knowledge gained from many tasks during meta-training to enable a model to quickly learn new tasks from few examples. Meta learning algorithm, or the learning method itself, such that the modified learner is better than the original learner at learning from additional experience. This paper explore introduction of Meta learning, how it works , the structure of literature survey of Meta learning, Meta Learning for few shot learning in specialized domains, evaluation metrics and benchmark datasets for meta learning and future direction and open problems in meta learning for few shot learning..

Introduction to Meta Learning

Meta Learning is known as "learning to learn". It is a branch of machine learning focuses on creating systems. Unlike traditional machine learning, which requires large datasets and extensive training for each new task, Meta learning aims to make models more flexible and data efficient. Importance of Meta learning in machine learning are rapid adaptation, improved generalization across diverse domains, and reduced training cost and support for continuous learning without forgetting previous knowledge.

Significance of Zero, one and few shot learning

These are the learning paradigms closely related to Meta learning that deal with limited data scenario:Zero shot learning (ZSL) refers to a model's ability to recognize or perform tasks for classes or scenarios it has never seen during training. The model must generalize its knowledge from known (seen) classes to unseen ones using semantic relationships or auxiliary information (e.g., attributes, textual descriptions, or embedding).Identifying rare or new objects or understanding unseen intents in NLP and detecting rare diseases not represented in training data are the example of zero shot learning.

One shot learning refers to a model learns to recognize a new class from only one example. Significance of one shot learning is efficient learning, ideal for limited data domain like facial recognition, security and personalized AI assistants and encourages transfer learning and metric learning. Real application of one shot learning is such as FACE ID systems, signature verification and character recognition

Few shot learning(FSL) refers to a model learns new classes or tasks from a very small number of labeled examples (typically 2-100).Significance of few shot learning is to bridges the gap between traditional supervised learning and one shot learning. It reduces data collection cost and time. Applications of few shot learning are medical diagnosis, robotics and NLP tasks like text classification with limited examples. Following table 1 shows various aspect such as data requirement, goal, key advantage and real application of zero shot, one shot and few shot learning.

Table 1: Different aspect of zero, one and few shot learning

Aspect	**Zero-Shot**	**One-Shot**	**Few-Shot**
Data Requirement	No examples	One example	Few examples
Goal	Generalize to unseen classes	Learn from one example	Learn efficiently from few examples
Key Advantage	Eliminates labelling	Fast adaptation	Reduces training data need
Real-World Use	NLP, vision, recommendation	Face recognition	Healthcare, robotics, NLP

Working of Meta learning

Training models to quickly adapt to new tasks with minimal data is the focus of a machine learning paradigm known as "meta-learning," or "learning to learn." In order to help models quickly adapt to new, untested tasks using a limited amount of task-specific data, meta-learning aims to enable models to generalize learning experiences across different tasks

Two primary phases are involved in the typical meta-learning workflow:

Meta – Learning Tasks:

Exposure to a range of tasks, each with its own set of parameters or characteristics, is part of the meta-training phase.

Model Training: Many tasks are used to train a base model, also known as a learner. The purpose of this model is to represent shared knowledge or common patterns among various tasks.

Adaption: With few examples, the model is trained to quickly adjust its parameters to new tasks.

Meta - Testing (Adaption)

New Task: The model is given a brand-new task during the meta-testing stage that it was not exposed to during training.

Few Shots: With only a small amount of data, the model is modified for the new task (few-shot learning). In order to make this adaptation, the model's parameters are frequently updated using the examples from the new task.

Generalization: Meta-learning efficacy is evaluated by looking at how well the model quickly generalizes to the new task. Subsequent paragraphs, however, are indented.

We can summarize the meta-learning workflow generally follows these steps.A meta-learner is trained on multiple different tasks, each with its own (often smaller) dataset. Through this training, it extracts patterns about what works best for efficient learning in various scenarios. When the meta-learner encounters a new task, it can quickly adapt and perform well, often with very little new data.

Taxonomy of Meta learning approaches

Model-Based Meta-Learning

Model-based meta-learning approaches employ specialized architectures designed to rapidly adapt to new tasks by encoding prior knowledge within their internal states or parameters. These models effectively function as adaptive learning systems that can generalize quickly with limited data, often by incorporating external memory or dynamic networks that facilitate fast learning. A prominent example is the use of Neural Turing Machines (NTMs) or other recurrent architectures with augmented external memory, which store task-relevant information that the model can retrieve and utilize during adaptation [7].

Memory-Augmented Matching Networks exemplify this approach by integrating such memory mechanisms to enhance one-shot and few-shot learning performance, achieving improved accuracy on benchmark datasets such as Omniglot and mini-Image Net. These architectures enable the network to simulate human-like rapid learning by recalling similar past experiences, thus enabling the model to efficiently infer new classes from scant data [7]. This model-based paradigm contrasts with static architectures by incorporating dynamic representations that evolve with new input, offering a flexible framework for handling diverse few-shot learning tasks.

Model-based meta-learning emphasizes the importance of adaptable model structures capable of encoding and efficiently accessing information relevant across tasks, supporting rapid generalization in scenarios where task boundaries and distributions vary significantly.

Metric-Based Meta-Learning

Metric-based meta-learning centers on learning similarity functions or embedding spaces that facilitate rapid comparison and classification of examples with few instances. Instead of directly adapting model parameters, these methods focus on learning a distance metric that generalizes well across tasks, allowing the system to classify new examples based on their proximity to known examples in the learned space.

Typical frameworks in this category include Matching Networks, which map queries and support examples to an embedding space where classification reduces to a weighted nearest neighbor search [8]. Relation Networks build on this idea by learning a deep non-linear distance function between query and support set instances, improving the flexibility and robustness of metric comparisons [8]. Prototypical Networks further simplify this concept by representing each class with a prototype—an average embedding computed over support examples—thereby facilitating efficient classification through metric distances [9].

These models have demonstrated strong performance on various few-shot benchmarks, outperforming traditional classifiers trained from scratch. Their reliance on learned metrics rather than parameter adaptation makes them particularly well-suited when rapid task adaptation is required without expensive fine-tuning. Moreover, metric-based methods have been effectively applied in specialized domains such as drug discovery, where data scarcity is acute, highlighting their utility in handling diverse real-world few-shot learning challenges [10].

Optimization-Based Meta-Learning

Optimization-based meta-learning adopts a fundamentally different strategy by focusing on learning how to optimize model parameters efficiently for new tasks. Frameworks like Model-Agnostic Meta-Learning (MAML) epitomize this category by learning an initialization of model parameters that can be rapidly adapted to new tasks with a few gradient updates [11]. This inner-outer loop optimization framework enables efficient transfer of knowledge through learned optimization trajectories, translating to faster convergence and improved performance on few-shot tasks.

Variants of MAML have addressed its computational and scalability challenges, including first-order approximations that avoid expensive second-order derivatives while maintaining competitive performance [12]. Further developments incorporate uncertainty-aware mechanisms through Bayesian formulations and variational inference, enhancing the model's robustness and calibration on unseen tasks [13]. These optimization-based approaches provide a flexible and theoretically grounded method for few-shot learning, capable of adapting arbitrary neural architectures within meta-learning frameworks.

Optimization-based meta-learning thus combines algorithmic innovation with practical efficacy, opening pathways for efficient model adaptation in domains ranging from image classification to sequential predictions, as it directly optimizes for rapid learning dynamics rather than just static representations.

Zero-Shot Learning and Meta-Learning Integration

Techniques for Zero-Shot Learning in Meta-Learning Context

Zero-shot learning (ZSL) represents a challenging frontier, demanding models to recognize classes or perform tasks without any direct labeled examples. Integrating meta-learning with zero-shot paradigms involves leveraging semantic embedding, auxiliary information, or multimodal data to bridge the gap between seen and unseen classes. For instance, learning a cross-modal mapping between semantic concepts and image features enables models to infer class characteristics based on attributes or textual descriptions [14]. In this context, meta-learning serves to enhance the generalization capability of such models by training them to adapt to novel modalities or semantic spaces efficiently, even in the absence of training samples from target classes.

More advanced approaches exploit multi-level meta-learning frameworks, incorporating deep kernel learning methods such as Deep Kernel Transfer (DKT). DKT implements a Bayesian treatment in the inner optimization loop to transfer kernels across tasks, enabling uncertainty quantification and robust zero-shot performance in classification and regression [15]. These methods alleviate reliance on task-specific parameter estimation, streamlining adaptation to unseen classes.

Meta-learning thus acts as a scaffold for zero-shot models to acquire the capacity for rapid generalization by optimizing representations or transfer mechanisms that exploit semantic similarity, paving the way for scalable zero-shot classification in practical applications where labeled data is prohibitive to obtain.

Meta-Meta Learning and Zero-Shot Classification

An emerging concept relevant to zero-shot learning is meta-meta learning. This involves learning not just to learn, but learning how to combine multiple meta-learners, each specialized in a particular type of learning problem. Such ensemble frameworks build meta-meta classifiers that determine, for any new task, the optimal combination of biased low-variance learners to achieve the best classification with minimal data [16]. This strategy proves particularly effective in zero-shot or extremely low-data settings since it leverages collective expertise without explicit task-specific training data.

Meta-meta classifiers can dynamically select, weight, and combine strategies tailored to task characteristics, outperforming traditional meta-learning or ensembling alone. This flexibility is crucial when tasks vary widely or when direct supervision is scarce, as in zero-shot classification.

Furthermore, advancements incorporate natural language guidance and generative latent space models from image synthesis to generate task-adapted neural network weights in a zero-shot manner, enhancing zero-shot learning's breadth and accuracy [17]. This creative intersection between meta-learning and generative models opens new directions for handling unseen tasks seamlessly.

Zero-Shot Learning Applications across Domains

Zero-shot learning's applicability spans diverse domains, particularly where domain shifts and modality heterogeneity present significant obstacles. Cross-lingual natural language understanding exemplifies such a domain, where meta-learning facilitates training on high-resource languages with minimal or zero data from target low-resource languages [18]. Through learning what knowledge to share and how to select beneficial instances, models achieve improved performance in zero-shot and few-shot cross-lingual tasks such as natural language inference and question answering.

In automated machine learning (AutoML), meta-learning frameworks extend to zero-shot pipeline and hyperparameter selection by learning surrogate models that rank and select optimal deep learning pipelines based on simple meta-features describing new datasets, enhancing efficiency under resource constraints [19]. Similarly, zero-shot summarization methods benefit from meta-learning to generalize across unseen document domains, especially when integrating multimodal or external knowledge sources [20].

Robotics also employs zero-shot meta-learning to adapt control policies or perception models to new tasks or environments without retraining, relying on learned meta-mappings and functional programming-inspired architectures to enable zero-shot task remapping [21]. These applications underscore the power of meta-learning to enable zero-shot capabilities across fields marked by rapid change, limited data, or multimodality.

One-Shot Learning Strategies via Meta-Learning

Episodic Training and Task Simulation

One-shot learning's hallmark is requiring models to classify or perform tasks based on just a single labeled example. A prevalent training technique involves episodic training, where the meta-learner is repeatedly trained on simulated tasks or episodes that mimic one-shot scenarios. This procedure conditions the model to acquire representations and learning dynamics attuned to rapid generalization [5]. By structuring training into episodes that closely replicate the sparse data conditions of testing, the model's ability to quickly adapt to new tasks is promoted.

Episodic training facilitates learning task-level inductive biases that transcend specific class boundaries, preparing the model for diverse one-shot classification challenges. This approach encourages extraction of transferable features and robust decision boundaries, helping mitigate model overfitting or bias issues associated with limited data.

Through episodic simulations, meta-learning frameworks develop a nuanced understanding of task variations, effectively preparing them for real-world settings where truly one-shot learning is demanded.

One-Shot Learning Architectures

Architectural innovations play a crucial role in enabling effective one-shot learning. Siamese Networks pioneered this area by learning a function that maps pairs of inputs to a similarity score, reducing classification to a verification problem by comparing a query instance with a single known example [8]. Relation Networks extend this principle by learning a deep non-linear metric for comparing pairs of images, trained end-to-end to infer relations and achieve improved classification accuracy within the one-shot realm [8].

Meta-meta classifiers further enhance one-shot approaches by combining multiple learners to tackle varied test problems, capitalizing on ensemble diversity [16]. These architectures focus on learning notions of similarity or relations rather than explicit class boundaries, benefiting from flexibility and data efficiency.

Notably, applications in robotics push the boundaries of one-shot learning architectures by enabling robots to imitate human actions from a single demonstration. Models trained via meta-learning learn priors from multiple tasks and contextually adapt to perform new manipulation skills based on a single video demonstration, overcoming domain shifts and embodiment differences [22]. This fusion of architectural design and meta-learning exemplifies the synergy needed to attain practical one-shot learning in complex settings.

Robotics and Imitation Learning from One Demonstration

Robotics presents an ideal application for one-shot learning via meta-learning, bridging perception, action, and domain adaptation challenges. Recent research has demonstrated systems capable of learning new manipulation skills by observing a single human demonstration video, even in the presence of significant domain discrepancies such as changes in perspective, environment, or embodiment between human and robot [22].

Meta-learning allows robots to build a prior from multiple previous related tasks, forming an adaptable knowledge base enabling fast learning of new ones. Experimental results on robotic arms like PR2 and Sawyer demonstrate successful rapid learning to perform pick-and-place, pushing, and placement tasks with just one human-provided video [23]. Such capabilities mark a significant advancement over traditional learning methods dependent on extensive, task-specific training.

The incorporation of domain-adaptive meta-learning in robotics thus paves the way for flexible, efficient skill acquisition resembling human learning capabilities, crucial for real-world deployment of autonomous systems.

Few-Shot Learning with Meta-Learning: Core Techniques

Data Augmentation and Embedding Strategies

One avenue to enhance few-shot learning involves leveraging data augmentation and robust embedding learning to compensate for scarce labeled samples. Data augmentation artificially increases dataset size and diversity by applying transformations or generative models to existing samples, helping prevent overfitting and improving generalization [4]. Embedding strategies focus on learning feature representations that capture transferable semantics, enabling models to identify and generalize patterns across tasks even with limited data.

Such techniques have been effectively employed in computer vision tasks and drug discovery applications, where learning meaningful embeddings allows models like Prototypical Networks or Matching Networks to perform reliably despite data constraints [10]. Enriching embeddings with intermediate representations or graph-based features can further enhance model robustness in few-shot scenarios.

Collectively, data augmentation coupled with thoughtfully learned embeddings lays a strong foundation for successful few-shot learning by amplifying available information and facilitating rapid knowledge transfer.

Optimization-Based Few-Shot Learning (e.g., MAML and Variants)

Optimization-based methods, particularly Model-Agnostic Meta-Learning (MAML), form a cornerstone of few-shot learning approaches. MAML learns an initialization of network parameters suitable for rapid adaptation through a small number of gradient steps on novel tasks, achieving superior generalization to unseen tasks [11]. Despite its popularity, MAML faces challenges relating to computational complexity due to second-order gradient calculations.

Addressing this, first-order approximations like TA-Reptile eliminate second-order terms, maintaining competitive performance while reducing computational costs substantially [12]. Moreover, Bayesian meta-learning variants employ gradient-based variational inference to model uncertainty in parameter estimates, leading to improved calibration and robustness in classification and regression tasks [13].

These optimization-centric frameworks offer a principled mechanism for fast learning in diverse domains, supporting adaptation from limited data with versatility and efficiency.

Metric Learning and Task-Adaptive Projections

Complementing optimization-based methods, metric learning approaches enhance few-shot learning by focusing on distinguishing inter-class differences through learned similarity functions. Task-adaptive projection models like TapNets project embedded features into task-specific subspaces, conditioning the model dynamically to particular few-shot learning scenarios and improving discrimination [9].

Prototypical networks aggregate support examples to form representative prototypes serving as centers of each class, facilitating efficient nearest-neighbor classification in the embedding space. The ability to learn these metrics and projections episodically enables models to generalize better across varied tasks [9]. Relation Networks extend the flexibility by learning complex similarity functions, enabling refined class discrimination especially under limited data.

Together, metric learning and task-adaptive projections form a robust strategy for few-shot classification, balancing flexibility and computational efficiency through learned comparison mechanisms.

Table 2 Different Categories of Learning along with key methods , challenges and their applications

Category	Core Idea	Key Methods / Models	Key Papers / References	Challenges	Applications
Few-Shot Learning	Learn from a small number of examples (2–100)	MAML, Prototypical Networks, Matching Networks, Relation Networks, Memory-Augmented Neural Networks (MANNs), Reptile	Finn et al. (2017), Snell et al. (2017), Vinyals et al. (2016), Santoro et al. (2016), Nichol et al. (2018)	Overfitting, high intra-class variance, poor generalization	Medical diagnosis, robotics, text classification
One-Shot Learning	Learn from just one labeled example per class	Siamese Networks, Matching Networks, NTMs, Data Augmentation, Memory-based models	Koch et al. (2015), Vinyals et al. (2016), Graves et al. (2014)	Generalization from single example, feature robustness	Face recognition, signature verification, character recognition
Zero-Shot Learning	Learn to classify unseen classes using side information	Semantic Embeddings (Word2Vec, GloVe, FastText), Generative Models (GANs, VAEs), Graph-based ZSL	Xian et al. (2017), Mikolov et al. (2013), Goodfellow et al. (2014), Socher et al. (2013), CVAE-ZSL	No training data for unseen classes, domain shift	Object recognition, NLP tasks, machine translation
Optimizatio n-Based Meta-Learning	Learn rapid model adaptation by optimizing initializatio n	MAML, Reptile, Meta-SGD	Finn et al. (2017), Nichol et al. (2018), Li et al. (2017)	Task adaptability, computational cost	Few-shot image classification, reinforcemen t learning
Metric-Based Meta-Learning	Learn similarity/d istance metrics for comparison	Siamese Networks, Prototypical Networks, Relation Networks	Koch et al. (2015), Snell et al. (2017), Sung et al. (2018)	Designing task-specific metrics, scalability	One-shot facial, medical imaging, handwriting recognition
Model-Based Meta-Learning	Rapid learning using memory/qu ery mechanism s	MANNs, NTMs, Meta-LSTM	Santoro et al. (2016), Graves et al. (2014), Ravi & Larochelle (2017)	Memory management, fast adaptation	NLP models, language translation, reinforcemen t learning

Bayesian Meta-Learning	Probabilistic generalization, uncertainty estimation	BMAML, CNAPs, PACOH	Yoon et al. (2018), Gordon et al. (2019), Rothfuss et al. (2021)	Modeling uncertainty, robustness	Few-shot NLP, medical AI, self-driving
Meta-RL	Fast adaptation in reinforcement environments	RL^2, VariBAD, MAML for RL	Duan et al. (2016), Zintgraf et al. (2019), Finn et al. (2017)[1]	Policy transfer, sample efficiency	Robotics, game AI, autonomous control
Neural Architecture Search&AutoML	Automatic search for effective network structures	ENAS, DARTS, AutoML-Zero	Pham et al. (2018), Liu et al. (2019), Real et al. (2020)	Search efficiency, integration with meta-learning	Model design, hyperparameter tuning, optimization
Gradient-Free Meta-Learning	Meta-learning without gradients or backpropagation	Evolutionary algorithms, black-box meta-learners	Lee et al. (2018), Real et al. (2020)	Sample inefficiency, convergence speed	Model adaptation, hyperparameter search
Domain Adaptation & Transfer Learning	Cross-domain generalization	Few-shot domain adaptation, meta-transfer learning	Cai et al. (2021), Tsai et al. (2020)	Quick adaptation, knowledge transfer	Cross-lingual NLP, healthcare diagnostics, cross-domain image recognition

The Structure of Literature Survey

This is literature survey section consist of different researcher overview regarding their opinion about Meta learning subdomains, Meta learning approaches ,challenges limitation and result observations.

Meta Learning is a process that helps models learns new and unseen tasks on their own. Figure 1 depicts the structure of literature survey of Meta learning. There are various subdomains of meta learning namely Optimization based Meta learning, Metric Based Meta Learning, Model Based Meta Learning, Bayesian Meta Learning, Few Shot One Shot Zero Shot Learning, Meta Reinforcement Learning, Meta Learning for Neural Architecture Search, Gradient Free Meta Learning, Meta Learning for Domain Adaption and Transfer Learning.

Researcher Chelsea Finn et al. [11][22] propose an algorithm for meta-learning that is model-agnostic, in the sense that it is compatible with any model trained with gradient descent and applicable to a variety of different learning problems, including classification, regression, and reinforcement learning. The goal of meta-learning is to train a model on a variety of learning tasks, such that it can solve new learning tasks using only a small number of training samples. In their approach, the parameters of the model are explicitly trained such that a small number of gradient steps with a small amount of training data from a new task will produce good generalization performance on that task. In effect, method trains the model to be easy to fine-tune. Authors demonstrate that this approach leads to state-of-the-art performance on two few-shot image classification benchmarks, produces good results on few-shot regression, and accelerates fine-tuning for policy gradient reinforcement learning with neural network policies.

The researcher Finn et al., Nichol et al. and Linn et al.[11][22] proposed the goal is to learn an optimization strategy by that allows models to quickly adapt to new tasks. Optimization-based methods, which are predominant among meta-learning solutions for parametric models, primarily focus on optimizing the initialization of the model parameters used during the training procedure. The rationale behind this approach is that effective initialization can expedite the adaptation of model parameters to new tasks with minimal optimization steps.

Examples of such schemes include the model agnostic meta-learning (MAML) algorithm and its variations [6], [7]. These methods may be extended to the design of other training algorithm hyper parameters such as the learning rate. Existing optimization-based methods that address model initialization can be further categorized into second-order and first-order algorithms, distinguished by their utilization of second-order derivatives and first-order gradient information. Additionally, modular meta-learning presents a distinct optimization-based approach that relies on the recombination of shared modules to address individual tasks[16]. The key methods are Model-Agnostic Meta-Learning (MAML) The initial model parameters are optimized so that a few gradient updates adapt it to a new task. Reptile are First-order approximation of MAML that simplifies the computation. Meta-SGD learns not just initialization but also update rules. There are application are few shot image classification, Adaptive robotics, Reinforcement learning.

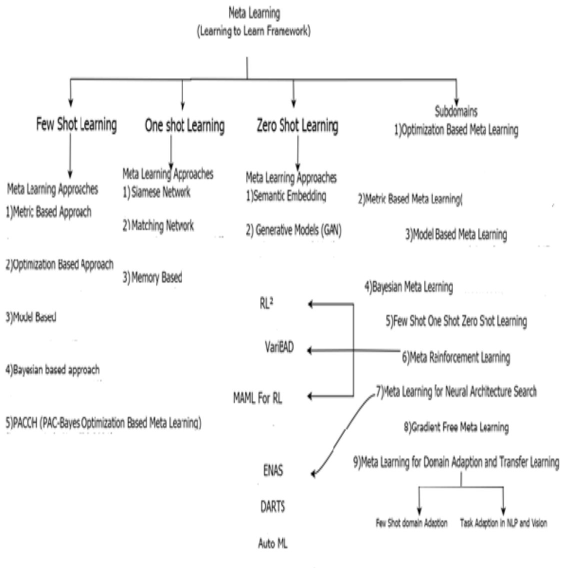

Figure 1 Literature Structure of Meta learning subdomains

Metric-based approaches, according to Snell et al. [60], work on the premise that tasks in the environment have a common feature representation mapping that makes it possible to measure how similar data points are. The objective is to acquire a similarity function enabling a model to categorize new instances based on their proximity to established instances. These methods make it easier to use nonparametric predictive models without having to train them on new tasks by meta-learning a similarity metric from data from many tasks. Siamese Network uses a distance metric (e.g., Euclidean distance) between embedding's of two images to determine similarity. Matching Networks Learns an embedding space and classifies new instances via a nearest-neighbor approach Notable examples of modern metric-based meta-learning methods include matching networks [3], prototypical networks [4], and relational networks [5].Prototypical Networks are Computes class prototypes and classifies new points based on distance. While aligned with the empirical Bayes methods seen in Gaussian Processes, the focus here is on collecting data from distinct tasks. However, due to their less frequent adoption in engineering problems, parametric models will be the primary focus of this monograph, leading to limited elaboration on metric-based meta-learning. Application based on metric meta learning are One shot facial recognition, handwriting recognition and Medical imaging.

Graves et al. [61] and Ravi and Larochelle et al. [62] proposed Model-based meta-learning involves creating models that have a memory part, which allows them learn rapidly from only a limited amount of examples. Examples of these kinds of models are Memory-Augmented Neural Networks (MANNs), Neural Turing Machines (NTMs), and meta-LSTM. Memory-Augmented Neural Networks use external memory, such as a differentiable memory module, to keep learned representations. Model-based approaches, on the other hand, involve optimizing a hypermodel that directly turns a task's training set into a model. There are numerous types of neural networks that are able to do this translation, including recurrent neural networks, convolutional neural networks, or hypernetworks. A straightforward example of model-based meta-learning uses the training set of a new task to refine a context vector that governs the operation of a model shared across tasks. Neural Turing Machines (NTMs) employ a memory matrix and learn read/write operations. Meta-LSTM uses LSTMs to derive update rules for neural networks instead of relying on backpropagation

Gordon et al. [23] provide thoughts on Bayesian Meta Learning. It uses probabilistic models to improve both the ability to estimate uncertainty and the ability to generalize. Bayesian MAML improves on MAML by learning a distribution over model parameters instead of using a fixed initialization. Conditional neural processes are used by CNAPs (Conditional Neural Adaptive Processes) to adapt to each task. PACOH (PAC-Bayes Optimization based Meta Learning) is a

Bayesian optimization framework that has been specifically designed for meta learning.

"One-shot Learning of Object Categories" by Fei-Fei, Fergus, and Perona [63] is one of the earliest approaches that use meta-learning for object recognition tasks. Instead of starting from the beginning, the authors utilized insights from previously acquired categories, regardless of how distinct these categories might be. This hypothesis was conducted via a Bayesian setting: The authors extracted "general knowledge" from previously learned categories and represented it in the form of a prior probability density function in the space of model parameters. Given a training set, regardless of how small, the authors updated this knowledge and produced a posterior density that could be used for object recognition. Their experiments showed that this was a productive approach, and that some useful information about categories could have been obtained from a few training examples.

"One-shot Learning of Simple Visual Concepts" by Lake, Salakhutdinov, Gross, and Tenenbaum [24] works in the domain of handwritten characters, an ideal setting for studying one-shot learning at the interface of human and machine learning. It presents a Hierachical Baysesian Model that learns visual concepts from a single example by inferring a probabilistic program. Handwritten characters contain a rich internal part structure of penstrokes, providing a good a priori reason to explore a parts-based approach to representation learning. The authors propose a new model of character learning based on the induction of probabilistic part-based representations. The model's approach, based on compositionality and causality, allows it to generate novel examples and perform one shot classification with human level performance, outperforming deep learning models on Omniglot dataset. Given an example image of a new character type, the model infers a sequence of latent strokes that best explains the pixels in the image by drawing on a broad stroke vocabulary abstracted from many previous characters. This stroke-based representation guides the generalization of new examples of the concept. This work aimed to bridge the gap between human ability to quickly learn concepts from minimal data and data-intensive requirements of traditional machine learning algorithms.

"One-shot Learning with a Hierarchical Nonparametric Bayesian Model" by Salakhutdinov, Tenenbaum, and Torralba [65] leverages higher-order knowledge abstracted from previously learned categories to estimate the new category's prototype as well as an appropriate similarity metric from just one example. These estimates are also improved, as more examples are observed. As illustrated in figure 2, consider how human learners seeing one example of an unfamiliar animal, such as a "wildebeest," can draw on experience with many examples of "horse," "cows," "sheep," etc. These similar categories have similar prototypes and share similar variations in their feature space representations. If we can identify the new example of "wildebeest" as belonging to this "animal" super-category, we can transfer an appropriate similarity metric and thereby generate informatively even from a single example. The algorithm that the authors used is a general-purpose hierarchical Bayesian model that depends minimally on domain-specific representations but instead learns to perform one-shot learning by finding more intelligent representations tuned to specific subdomains of a task.

"One-shot Learning by Inverting A Compositional Causal Process" by Lake, Salakhutdinov, and Tenenbaum [66] tackles one-shot learning via a computational approach called Hierarchical Bayesian Program Learning that utilizes the principles of composition and causality to build a probabilistic generative model of handwritten characters. This is compositional because characters are represented as stochastic motor programs where the primitive structure is shared and reused across characters at multiple levels, including strokes and sub-strokes. Given the raw pixels, the model searches for a "structural description" to explain the image by freely combining these elementary parts and their spatial relations. This is causal because strokes are not modeled at the level of muscle movements, but are abstract enough to be completed by higher-order action. The model was evaluated on the Omniglot dataset, a challenging benchmark for one-shot classification and generation tasks, achieving human-level performance. The model's ability to produce human-like performance on tasks beyond simple classification, such as generating new examples of a learned concept, demonstrates a deeper conceptual understanding. In essence, the paper shows that by modeling the underlying compositional and causal processes that create visual concepts, a system can learn to recognize and generate new visual classes from very few examples, mimicking human-like one-shot learning abilities.

Few Shot One Shot Zero Shot Learning
Few shot, one shot and zero shot learning enable models to generalize from limited or no train examples. Few Shot Learning methods are prototypical network, MAML and Meta-LSTM.One shot Learning methods are Siamese Networks and Matching network. Whereas zero shot learning methods are Semantic embeddings (Word2Vec,GloVe,FastText) and Generative Models which include Generative adversarial Network and Variational Autoencoder(VAE). Applications are Image classification with limited data, zero-shot machine translation and few shot medical diagnosis.

Meta Reinforcement Learning
RL2 (Reinforcement Learning Squared) uses an RNN to encode task information for quick adaptation.VariBAD (Variational Bayes Adaptive RL) is Bayesian inference-based method for adaptive RL.Model Agnostic Meta Learning (MAML) are useful in solving reinforcement learning problems.

Meta Learning for Neural Architecture Search
Meta Learning for neural architecture search (NAS) is to automate the search for optimal neural network architectures. Key Methods include ENAS (Efficient NAS) uses reinforcement learning for efficient architecture search. DARTS (Differentiable Architecture Search) uses gradient-based optimization for NAS. AutoML-Zero is Evolutionary-based NAS framework. Application examples includes Automated deep learning , model design Hyper parameter tuning and Neural network optimization

Gradient Free Meta Learning
Gradient Free Meta Learning techniques is meta learning techniques that do not rely on gradient based optimization. The key methods used in Gradient free meta learning include black box meta learning and evolutionary algorithms for meta learning. Applications are evolutionary based model adaptation and Hyperparameter search.

Meta Learning for Domain Adaption and Transfer Learning
It enable models to transfer knowledge across different domains with minimal retraining. The key methods are few shot domain adaptation and Meta learning for cross domain tasks. Applications are Cross-lingual NLP models, Cross-domain image recognition, AI-assisted healthcare diagnostics

Table 2: Meta-Learning for Few-Shot Learning in Specialized Domains

Paper / Author(s)	Meta-Learning Approach	Challenges	Limitations	Results & Observations
Gharoun et al. (Meta-Learning Approaches Survey),2024	Comprehensive review of meta-learning approaches in few-shot	- Scalability and efficiency - Generalization across diverse tasks	- Full survey; limitations context-dependent	Reviews state-of-the-art meta-learning paradigms and provides insights into recent advances and research directions in few-shot learning
Verma et al. (Meta-Learning for Generalized Zero-Shot Learning),2020	Model-Agnostic Meta-Learning (MAML) integrated with Wasserstein GAN	- Training generative models with few seen class samples - Generalizing to unseen classes - Bias towards seen classes in GZSL	- GAN training stability - Mode collapse on fine-grained datasets - Requires attribute vectors for classes	Significant improvement in zero-shot (ZSL) and generalized zero-shot learning (GZSL) tasks with few-shot seen samples; relative improvements up to 27.9% over state-of-the-art on standard benchmarks (CUB, AWA2, aPY). Meta-learned GANs generate high-quality samples enabling better unseen class generalization. Proposed zero-shot task distribution enhances performance[57][58]

Vinyals et al. (Matching Networks for One-Shot Learning),2016	Metric-based Meta-Learning with attention and memory	- Overfitting with few examples - Need to rapidly adapt to new classes - Defining appropriate similarity metric	- Requires episodic training - High computational cost for large support sets	Achieved state-of-the-art one-shot accuracy on Omniglot and ImageNet; episodic training mimics test conditions improving generalization. Demonstrated effectiveness of attention-weighted nearest neighbor classification in embedding space [33]
Snell et al. (Prototypical Networks for Few-Shot Learning),2017	Metric-based: Learning class prototypes and distances	- Limited labeled examples per class - Handling class variability	- May not handle complex task distributions	Robust few-shot classification via embedding with learned prototypes; simple and effective, scalable. Shows strong performance on benchmark datasets like miniImageNet [60]
Finn et al. (MAML: Model-Agnostic Meta-Learning),2017	Optimization-based Meta-Learning	- Fast adaptation to new tasks with few samples - Computational demands of nested gradients	- Sensitive to hyperparameters - Requires differentiable models	Provides a model initialization enabling quick fine-tuning; broadly applicable across domains (RL, vision). Forms basis of ZSML in zero-shot learning[22]
Other Generative Models in Zero-Shot Learning (f-CLSWGAN, CVAE-ZSL),2017-2018	Generative Adversarial Networks and Variational Autoencoders for Synthetic Data	- Training instability of GANs - Data imbalance - Poor quality on unseen class generation	- Often require large seen class data - Hard to ensure sample diversity	Generative models alleviate bias in GZSL by synthesizing unseen class samples; inferior when training data for seen classes is scarce

Meta-Learning For Few-Shot Learning In Specialized Domains

Cyberspace Security and Intrusion Detection

In the cybersecurity domain, signaling the high stakes of identifying zero-day and emerging attacks under data scarcity, meta-learning offers promising solutions. Traditional supervised models require large-scale labeled logs, which are often unavailable, making few-shot and zero-shot learning vital for robust intrusion detection [3]. Meta-learning frameworks in this realm employ metric-based and optimization-based techniques to detect anomalies and classify attack types with limited samples.

For instance, deep neural network architectures and meta-learning strategies have been proposed to distinguish network traffic patterns, achieving high detection rates even when trained with minimal malicious samples [24]. Furthermore, continual few-shot learning methods enable intrusion detection models to adapt online to new attack types without forgetting previous knowledge, a critical requirement for evolving security landscapes [25]. By enabling fast adaptation and generalization across diverse and changing threats, meta-learning strengthens the resilience and responsiveness of cybersecurity defenses.

Medical Imaging and Bioacoustics Applications

Medical imaging frequently encounters the challenge of limited annotated data, compounded by diversity across imaging sites and patient populations. Site-agnostic meta-learning addresses this by learning generalized initializations that adapt efficiently to new sites with few examples, improving classification accuracy in diagnosing conditions such as autism spectrum disorder [26]. These methods handle heterogeneity in imaging protocols and patient demographics, yielding robust clinical models from scarce data.

Bioacoustic event detection similarly benefits from few-shot learning methodologies. Instead of deploying data-hungry deep learning, some approaches leverage classical machine learning techniques augmented by meta-learning to classify rare acoustic events in wildlife monitoring, balancing accuracy with computational efficiency [27]. Moreover, the integration of meta-learning with Neural Architecture Search (NAS) in medical image analysis optimizes network structures for few-shot tasks, delivering improved segmentation and classification performance while reducing reliance on large annotated datasets [28].

These advances demonstrate meta-learning's transformative impact on healthcare and ecological monitoring, enabling effective analysis despite sparse data.

Financial Forecasting and Urban Infrastructure Monitoring

Financial markets are characterized by volatility and limited data in novel conditions, challenging traditional predictive models reliant on large historical datasets. Meta-learning and few-shot learning frameworks, incorporating model-agnostic meta-learning and Siamese networks, enable financial models to forecast price movements and volatility under unfamiliar market states without extensive retraining, facilitating adaptive and resilient financial modeling [6].

Urban infrastructure monitoring, particularly through distributed acoustic sensing (DAS), also leverages meta-learning for few-shot classification of diverse event types with minimal labeled samples. By exploring multiple data preprocessing techniques and embedding networks trained on meta-datasets, these frameworks achieve precise classification while accommodating varied sensor modalities [29]. This capability is particularly valuable for detecting anomalies or events in complex urban environments with limited annotated data, underpinning smart city applications.

Thus, meta-learning enhances the effectiveness and adaptability of models in both financial and urban infrastructure domains where data scarcity and domain heterogeneity pose significant challenges.

Challenges and Limitations in Meta-Learning for Few-Shot Settings

1. Data Distribution Shifts and Domain Adaptation

A fundamental challenge in meta-learning arises from distributional shifts between training and deployment domains. Cross-domain generalization remains difficult due to differing feature distributions, class semantics, or data modalities, which can substantially degrade few-shot learning performance. Addressing this requires domain-adaptive meta-learning approaches that explicitly model or compensate for such shifts, facilitating transferability [30]. For example, site-agnostic meta-learning methods in medical imaging explicitly optimize for robust adaptation across heterogeneous imaging centers, preserving performance despite domain variations [26]. Nonetheless, fully overcoming domain mismatch remains an open problem, especially in complex multimodal and noisy environments.

2. Computational Complexity and Scalability

Many meta-learning techniques, especially those based on optimization (e.g., MAML), involve costly second-order derivative computations, limiting scalability and applicability in resource-constrained settings. This computational burden limits real-time adaptation and deployment on edge devices.

To mitigate this, approaches such as first-order gradient approximations reduce overhead with minimal performance loss [12]. Additionally, hardware-aware meta-learning algorithms incorporate quantization and other hardware constraints upfront, optimizing learning efficiency and convergence speed without compromising accuracy [31]. These developments are crucial for meta-learning's transition from research prototypes to practical, scalable solutions.

3. Overfitting and Model Bias in Meta-Training

Overfitting to meta-training tasks is another significant concern. Meta-learners often develop biases toward tasks seen during training, limiting their adaptability to novel or unrelated tasks. Investigations into MAML reveal that feature reuse rather than rapid feature adaptation dominates its generalization, prompting the development of simplified algorithms and regularization methods to balance these effects [11].

Entropy-based methods and task-agnostic meta-learning strategies have also been proposed to combat model bias and overfitting, enhancing the generalization across diverse unseen tasks [12]. Despite these advances, understanding and controlling meta-learner biases remains a key research direction to ensure broad applicability

Evaluation Metrics and Benchmark Datasets For Meta-Learning

Popular Benchmarks for Few-Shot and Zero-Shot Learning

Benchmark datasets such as Omniglot, mini-ImageNet, and Tiered-ImageNet have become standard for evaluating meta-learning performance, providing controlled settings for few-shot classification with well-defined splits of seen and unseen classes [5]. Omniglot offers a character recognition challenge with many classes but few examples per class, ideal for testing generalization. Mini-ImageNet and

Tiered-ImageNet provide more complex natural image datasets with increasing difficulty and diversity.
While widely adopted, limitations exist, including domain specificity and the challenge of replicating real-world heterogeneous data distributions. Domain-specific datasets in cybersecurity, medical imaging, and others are increasingly used to complement these benchmarks, highlighting diverse application challenges.

Common Evaluation Metrics
Evaluation metrics in few-shot learning typically include accuracy, precision, recall, and F1-score, especially for classification tasks [33]. Recent advances propose novel metrics adapted to specific domains. For instance, mean average log percentage error (MALPE) has been introduced in forecasting under limited data as a robust alternative to Mean Absolute Percentage Error (MAPE), mitigating bias and improving fairness in evaluation [34].
Such metrics ensure that performance assessments account for the unique difficulties inherent in few-shot setups, including class imbalance, variability, and sample scarcity.

Comparative Performance of Meta-Learning Approaches
Comparative studies show that no single meta-learning approach universally dominates; rather, relative performance depends on data type, task difficulty, and resource availability. Metric-based methods often provide efficient classification with good generalization, especially when paired with embedding augmentation [10]. Optimization-based models excel in rapid parameter adaptation but may incur computational costs, balanced by first-order variants [35].
Understanding these trade-offs is essential in selecting appropriate methods tailored to application constraints and objectives.

Future Directions and Open Problems In Meta-Learning For Few-Shot Learning
Enhancing Generalization and Robustness across Tasks
Future meta-learning research aims to enhance model generalization and robustness in the face of task heterogeneity and multimodal data distributions. Addressing complex, multimodal tasks requires methods that effectively integrate diverse input types and adapt to shifting task domains [36]. Expanding meta-learning's capacity to generalize across broader distributions and uncertainty conditions remains a critical goal [1].

Integration with Multimodal and Cross-Domain Learning
Leveraging multimodal information and semantic concepts promises to bridge gaps in zero- and few-shot learning by enriching representations and enabling cross-modal transfer [14]. Meta-learning models capable of learning shared latent spaces across modalities can better exploit the complementary nature of data sources, improving prediction accuracy and robustness in complex real-world settings [18].

Meta-Learning Under Resource Constraints and Real-World Deployment
Efficient meta-learning compatible with hardware and computational resource constraints is vital for real-world deployment. Advances in hardware-aware meta-learning accommodate quantized networks and optimize training under memory and speed limitations, facilitating few-shot learning at the edge [31]. Simultaneously, integrating meta-learning into scalable AI systems requires balancing adaptation speed, accuracy, and resource usage, a vital research frontier for making few-shot learning ubiquitous.

In conclusion, meta-learning has established itself as a versatile and powerful framework addressing the fundamental challenges of zero-, one-, and few-shot learning across various domains. By systematically learning how to learn, these approaches provide robust, scalable solutions that extend machine learning's applicability to situations marked by data scarcity and task variability. Ongoing advances in architectures, optimization techniques, domain adaptation, and resource-aware designs continue to expand its horizons, promising impactful applications in security, medicine, finance, and beyond.

References
[1] Harlow, H. F. (1949). The formation of learning sets. Psychological Review,56(1), 51–65. https://doi.org/10.1037/h0062474
[2] Sepp Hochreiter et.al "Learning to Learn Using Gradient Descent" 2001 Conference Paper DOI: 10.1007/3-540-44668-0_13 · Source: DBLP
[3] Lemke, C., Budka, M. & Gabrys, B. Metalearning: a survey of trends and technologies. Artif Intell Rev 44, 117–130 (2015). https://doi.org/10.1007/s10462-013-9406-y
[4] Hassan Gharoun,Fang chen et al "Meta-Learning Approaches for few shot learning:

A survey of recent adavances ACM Comput.Surv. 2024 DOI : 10.1145/3659943
[5] Gregory Koch, Gregory, Zemel, Richard, Salakhutdinov Ruslan "Siamese Neural Networks for one-shot Image Recognition Volume 12 2024 DOI : 10.1109/ACCESS.2023.3346.273
[6] Dash A Ye J Wang G "A Review of Generative Adversarial Networks (GANs) and Its Applications in a Wide Variety of Disciplines: From Medical to Remote Sensing" IEEE Access volume 12 p.p18330-18357 2024 doi : 10.1109/ACCESS.2023.3346273
[7] Gummadi, V. P. K. (2022). *MuleSoft API Manager: Comprehensive lifecycle management*. Journal of Information Systems Engineering and Management**,** 7(4), 1–9. https://doi.org/10.52710/cfs.886
[8] Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. Nature 323, 533–536 (1986). https://doi.org/10.1038/323533a0
[9] Hinton, Geoffrey & Osindero, Simon & Teh, Yee-Whye. (2006). A Fast Learning Algorithm for Deep Belief Nets. Neural Computation. 18. 1527-1554. 10.1162/neco.2006.18.7.1527.
[10] Anna Vettoruzzo et al. Advances and Challenges in Meta-Learning: A Technical Review IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 46, Issue 7 Pages 4763 – 4779 https://doi.org/10.1109/TPAMI.2024.3357847 Published: 01 July 2024
[11] Chelsea Finn, Pieter Abbeel, Sergey Levine "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks" ICML 2017arXiv :1703.03400.
[12] Schmidhuber, J. (1993). A 'Self-Referential' Weight Matrix. In: Gielen, S., Kappen, B. (eds) ICANN '93. ICANN 1993. Springer, London. https://doi.org/10.1007/978-1-4471-2063-6_107
[13] Li, Yitong and Murias, michael and Major, samantha and Dawson, geraldine and Dzirasa, Kafui and Carin, Lawrence and Carlson, David E "Targeting EEG/LFP Synchrony with Neural Nets" 2017 Advances in Neural Information Processing Systems ,Curran Associates, Inc Publisher Volume 30
[14] Yulong Ji and Kunjin Zou and Bin Zou "Mi-maml: classifying few-shot advanced malware using multi-improved model-agnostic meta-learning"
[15] Jaoquin Vanschoren "Meta Learning : A Survey" 2018 arXiv:1810.03548
[16] Alom, Morshed. (2024). Meta-Learning: Adaptive and Fast Learning Systems. Journal of Artificial Intelligence General science (JAIGS) ISSN:3006-4023. 2. 90-97. 10.60087/jaigs.v2i1.p97.
[17] Yan Duan and John Schulman and Xi Chen and Peter L. Bartlett and Ilya Sutskever and Pieter Abbeel 2016 RL^2 : Fast Reinforcement Learning via slow reinforcement learning arXiv:1611.02779
[18] Jaesik Yoon, Taesup Kim, Ousmane Dia, Sungwoong Kim,Yoshua Bengio Sungjin Ahn 2018 "Bayesian Model-Agnostic Meta-Learning" Neural Information Processing Systems (NeuroIPS 2019)
[19] Ramírez, J. G. C., & mafiqul Islam, M. (2024). Application of Artificial Intelligence in Practical
[20] Scenarios. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 2(1), 14 DOI:https://doi.org/10.60087/jaigs.v2i1.41
[21] Luisa M Zintgraf and Kyriacos Shiarlis and Vitaly Kurin and Katja Hofmann and Shimon Whiteson 2019 "Fast Context Adaptation via Meta-Learning" arXiv:1810.03642
[22] Chelsea Finn and Pieter Abbeel and Sergey Levine 2017 "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks" arXiv:1703.03400
[23] Chuan Guo, Jacob R. Gardner ,Yurong You , Andrew Gordon Wilson, Kilian Q. Weinberger 2019 "Simple Black-box Adversarial Attacks" arXiv:1905.07121v2
[24] Lake, B.M., Salakhutdinov, R., Gross, J., & Tenenbaum, J.B.et al."One shot learning of simple visual concepts". Cognitive Science, 33.,2011
[25] Kondapalli vinay kumar.vipin milind kamble "One shot face recognition" PCEMS,2023
[26] Virat kumar K. Kothari, Dr Sanjay M. Shah "Automated Facial Recognition in Older Photographs Using One-Shot Learning in Siamese Networks and Transfer Learning"Turkish journal of computer and mathematics Education TURCOMAT 2019 Vol. 10 No. 3 (2019); 1609-1621 ; 1309-4653
[27] Patel Yogesh & Tanwar Sudeep & Gupta, Rajesh & Bhattacharya, Pronaya & Davidson, Inno & Nyameko, Royi & Aluvala, Srinivas & Vimal, Vrince "Deepfake Generation and Detection Case Study and

Challenges" IEEE Access Volume 11 Dec 2023
[28] Xue Wanqi ,Wang Wei "One-Shot Image Classification by Learning to Restore Prototypes" Conference paper in AAAI 2020 arXiv:2005.01234
[29] Gui Jie, Sun Zhenan,Wen Yonggang et al. "A Review on Generative Adversarial Networks: Algorithms, Theory and Application 2020 arXiv:2001.06937
[30] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and Philip Yu. 2022. Generalizing to unseen domains: A survey on domain generalization. IEEE Transactions on Knowledge and Data Engineering (2022).
[31] Creswell A White T Dumoulin V Arulkumaran K Sengupta B Bharath A et. al. "Generative Adversarial Networks: An Overview" 2018 arXiv:1710.07035
[32] Y. Xie, Y. Fu, Y. Tai, Y. Cao, J. Zhu and C. Wang, "Learning to Memorize Feature Hallucination for One-Shot Image Generation," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 9120-9129, 2022 doi: 10.1109/CVPR52688.2022.00892
[33] Oriol Vinyals, Charles Blundell, Timothy Lillicrap,Koray Kavuckcuoglu, Daan Wierstra "Matching Networks for One shot Learning" Advances in Neural information Processing systems 29 (2016) arXiv:1606.04080
[34] Jiang Lu, Pinghua Gong, Jieping Ye, and Changshui Zhang. 2020. Learning from very few samples: A survey. arXiv preprint arXiv:2009.02653 (2020).
[35] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. 2020. Tent: Fully test-time adaptation by entropy minimization. arXiv preprint arXiv:2006.10726 (2020).
[36] Chelsea B. Finn. 2018. Learning to Learn with Gradients. University of California, Berkeley.
[37] Abolfazl Farahani, Sahar Voghoei, Khaled Rasheed, and Hamid R. Arabnia. 2021. A brief review of domain adaptation. Advances in Data Science and Information Engineering (2021), 877–894
[38] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. 2014. How transferable are features in deep neural networks? Advances in Neural Information Processing Systems 27 (2014)
[39] Sachin Ravi and Hugo Larochelle. 2016. Optimization as a model for few-shot learning. International Conference on Learning Representations (2016)
[40] Sebastian Ruder, Matthew E. Peters, Swabha Swayamdipta, and Thomas Wolf. 2019. Transfer learning in natural language processing. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials. 15–18.
[41] Mei Wang and Weihong Deng. 2018. Deep visual domain adaptation: A survey. Neurocomputing 312 (2018), 135–153
[42] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 22, 10 (2009), 1345–1359
[43] Manali Shaha and Meenakshi Pawar. 2018. Transfer learning for image classification. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 656–660.
[44] Hanzi Xu, Muhao Chen, Lifu Huang, Slobodan Vucetic, Wenpeng Yin "X-Shot: A Unified System to Handle Frequent, Few-shot and Zero-shot Learning Simultaneously in Classification (2024)" [doi]: https://dx.doi.org/10.48550/arxiv.2403.03863
[45] Shafin Rahman, Salman Khan, Fatih Porikli "A Unified approach for Conventional Zero-shot, Generalized Zero-shot and Few-shot Learning." (2018-07-30)
[46] Xiaoxu Li, Xiaochen Yang, Zhanyu Ma, Jing-Hao Xue "Deep Metric Learning for Few-Shot Image Classification: A Review of Recent Developments (2023)"Pattern Recognition , 138 , Article 109381. (2023)
[47] Mesay Samuel, Lars Schmidt-Thieme, DP Sharma, Abiot Sinamo, Abey Bruck "Offline Handwritten Amharic Character Recognition Using Few Shot learning "(2022)[doi]: https://dx.doi.org/10.48550/arxiv.2210.00275
[48] Archit Parnami, Minwoo Lee "Learning from Few Examples: A Summary of Approaches to Few Shot learning(2022)[doi]: https://dx.doi.org/10.48550/arxiv.2203.04291
[49] Jishnu Jaykumar P, Yu-Wei Chao, Yu Xiang "FewSOL: A Dataset for Few Shot Object Learning in Robotic Environments" (2022) arxiv 2207.03333
[50] Shivaank Agarwal, Ravindra Gudi, Paresh Saxena " One-Shot learning based classification for segregation of plastic waste" (2020) arxiv.2009.13953

[51] Leanne Nortje, Herman Kamper" Unsupervised vs. transfer learning for multimodal one shot matching of speech and images" (2020) [doi]: https://dx.doi.org/10.48550/ arxiv.2008.06258
[52] A.V. Uzhinskiy, G.A. Ososkov, P.V. Goncharov, A.V. Nechaevskiy, A.A. Smetanin"One shot learning with triplet loss for vegetation classification tasks" (2021-07-01)Компьютерная оптика,(Computer Optics) Vol 45, Iss 4, Pp 608-614 (2021)[doi]: https://doi.org/10.18287/24 12-6179-CO-856
[53] Arkabandhu Chowdhury, Dipak Chaudhari, Swarat Chaudhuri, Chris Jermaine "Meta-Meta Classification for One-shot Learning" (2020) [doi]: https://dx.doi.org/10.48550/ arxiv.2004. 08083
[54] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, Bernt Schiele "Meta-Transfer Learning for Few-Shot Learning" (2018)[doi]: https://dx.doi.org/ 10.48550/arxiv.1812.02391
[55] Jianyi Li, Guizhong Liu " Few-Shot Image Classification via Contrastive Self-Supervised Learning" (2020)[doi]: https://dx.doi.org/ 10.48550/arxiv.2008.09942
[56] Jishnu Jaykumar P, Yu-Wei Chao, Yu Xiang "FewSOL: A Dataset for Few-Shot Object Learning in Robotic Environments" (2022) [doi]: https://dx.doi.org/ 10.48550/arxiv. 2207. 03333
[57] Archit Parnami, Minwoo Lee " Learning from Few Examples : A Summary of Approaches to Few-shot Learning"(2022)DOI : 10.48550/arxiv.2203.04291
[58] Vinay Kumar Verma, Dhanajit Brahma, Piyush Rai "Meta-Learning for Generalized Zero-Shot Learning" Proceedings of the AAAI Conference on Artificial Intelligence, 34(04), 6062-6069. https://doi.org/10.1609/aaai.v34i04.6069
[59] Dhanajit Brahma, Vinay Kumar Verma and Piyush Rai "Hypernetworks for Continual Semi-Supervised Learning" 2021 arXiv:2110.01856v1
[60] Jake Snell, Kevin Swersky, Richard S Zemel "Prototypical networks for few-shot learning" Advances in Neural Information Processing Systems(NIPS) 30, pages 4077-4087 ,2017
[61] Alex Graves, Greg Wayne and Ivo Danihelka, "Neural Turning Machines" 2014 [doi] https://doi.org/10.48550/arXiv.1410.5401
[62] Sachin Ravi and Hugo Larochelle "Optimization as a model for few-shot learning" Conference paper at ICLR 2017
[63] Li Fei-Fei, R. Fergus and P. Perona, "One-shot learning of object categories," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 4, pp. 594-611, April 2006, doi: 10.1109/TPAMI.2006.79
[64] Salakhutdinov, R.; Tenenbaum, J.; Torralba, A. "One-shot learning with a hierarchical nonparametric Bayesian model" PMLR 195--206 ,2012
[65] Lake, Brenden M and Salakhutdinov, Russ R and Tenenbaum, Josh," One-shot learning by inverting a compositional causal process" NIPS2013 Vol 2