Uncertainty-Aware Prototypical Networks with Monte Carlo Dropout for Few-Shot Image Classification

Syeda Roohi Fatema
Department of Information Science and Engineering, Ramaiah Institute of Technology, Bengaluru, India
srfatema@msrit.edu (corresponding author)

Sumana Maradithaya
Department of Information Science and Engineering, Ramaiah Institute of Technology, Bengaluru, India
sumana.m@msrit.edu

Received: 31 August 2025 | Revised: 9 October 2025 and 15 October 2025 | Accepted: 18 October 2025

DOI: https://doi.org/10.48084/etasr.14448

ABSTRACT

Meta-learning is a transformative method for accelerating efficient task adaptation within significantly constrained data regimens. Its multidimensional applications expand to the domains of few-shot classification, reinforcement learning, and domain generalization. Deep learning architectures necessitate expansive datasets to accomplish robust generalization, whereas meta-learning equips computational models with an exceptional capacity for competent adaptability within data-constrained and heterogeneous learning scenarios. However, despite its considerable potential, present meta-learning methods are substantially hindered by a critical epistemic limitation, i.e., their systematic inability to provide uncertainty estimates. This insufficiency results in overconfident predictive outputs that invariably hinder the practical implementation of such cutting-edge computational frameworks in real-world scenarios. In response to these limitations, this study presents a pioneering architectural framework that integrates prototypical networks with Monte Carlo (MC) dropout. Prototypical networks, established for their exceptional efficiency in few-shot learning scenarios, is one of the prominent meta learning algorithms. The proposed method utilizes class prototypes within latent embedding spaces for each class to expedite robust task generalization. MC dropout provides a meticulous probabilistic mechanism for uncertainty quantification through stochastic representations. The proposed algorithm improves the model's generalization accuracy and prediction reliability using uncertainty estimation. Experiments on image classification tasks demonstrate that dropout regularization improves performance and reduces overfitting in metric-based meta-learning algorithms. This comprehensive paradigm establishes a significant meta-learning application by overcoming the limitations of adaptation and uncertainty quantification, with profound implications for critical real-world scenarios.

Keywords-meta learning; few-shot learning; Monte Carlo dropout; uncertainty quantification; image classification; artificial intelligence

I. INTRODUCTION

In the rapidly evolving world of artificial intelligence, the ability to learn quickly and adapt to new tasks with limited data has emerged as a fundamental challenge. Traditional deep learning algorithms perform well in large labeled datasets. However, their effectiveness often declines when faced with few training samples and different task distributions [1]. This limitation raises significant issues in real-world applications, where data collection can be computationally expensive, time-consuming, or fundamentally limited [2]. Meta-learning offers a promising way to address the limitations of deep learning methods by utilizing a bi-level optimization approach with an inner loop and an outer loop. The inner loop quickly adapts to individual tasks using a few gradient steps to train the model for that specific task. The outer loop then learns from similar tasks, adjusting the model's initial parameters so that those inner loop steps work accurately. Essentially, the outer loop learns how to learn, focusing on adaptability instead of just performance on one task. Meta-learning differs from standard machine learning algorithms as it seeks to find common learning techniques that can be transferred across various domains. Meta-models can quickly adapt to new tasks by authorizing models that mimic human flexibility without requiring extensive training [3].

Deep learning algorithms usually rely on large datasets and significant computational resources for learning specific tasks. At its core, meta-learning extends beyond past task-specific training scenarios, as meta-learning algorithms search for core learning mechanisms by enabling generalization from a minimal dataset [4]. In contrast, traditional deep learning models require large amounts of training data to create reliable representations. Meta-learning generally refers to an agent's ability to recognize and refine its learning procedure, thereby increasing its learning capacity. This method of learning from task-specific to learning-to-learn techniques represents a significant move toward more intelligent and adaptive systems.

However, current meta-learning approaches face a fundamental shortcoming, i.e., the lack of effective methods for measuring uncertainty levels. In general, meta-models generate stand-alone predictions but avoid providing users with information about their prediction accuracy [5]. High-stakes scenarios require an accurate understanding of model confidence, as prediction reliability is equally essential to prediction accuracy. Most scenarios require a large amount of training data, and their effectiveness is strongly dependent on carefully labeled datasets with specified features. Such dependencies limit their ability to generalize accurately under changing circumstances. Numerous research studies have addressed this respective issue, with a primary focus on unsupervised and semi-supervised model training [6].

Meta-learning empowers models to adapt to new tasks via leveraging prior knowledge, thereby addressing generalization challenges in few-shot learning. It involves optimizing algorithm selection and parameter adjustment to handle new and similar problems efficiently. Current methods suffer from the issue of unmeasured uncertainty levels, leading to overfitting models [7, 8]. Meta models usually output single predictions but do not inform the users of the level of confidence for the given prediction. Training a model on multiple datasets does not lead to domain adaptation, although its generalization ability is significantly affected by the data distribution. Transfer learning improves model performance on target domains with limited labeled data by leveraging knowledge from related source domains and has proven effective in several applications. However, its effectiveness often diminishes when fine-tuning with a small number of samples or when samples differ significantly from the original training data [9]. In addition, such approaches are prone to overfitting due to insufficient training data for effective model adaptation.

Monte Carlo (MC) dropout has emerged as a principal approach to estimate uncertainty in neural networks, treating dropout as an approximate Bayesian inference [10]. MC batch normalization, a method parallel to MC dropout, was presented as an approximation inference method for neural networks and a model-independent algorithm that adjusts dropout rates to mitigate overfitting in DNNs [11, 12]. Targeted dropout has been proposed to enhance robustness for feature pruning during the training of large networks with a self-reinforcing sparsity criterion [13, 14]. R-Drop [15] was proposed as a training approach to regularize dropout to achieve regularization between the output distributions of the two dropout sub-models [15]. Transformer-based vision tasks have also incorporated dropout as a layer-wise unfreezing enhanced dropout to reduce overfitting [16].

Several approaches have been proposed to enhance the training of deep neural networks for supervised learning, building on the dropout method; however, their integration with metric-based meta-learning frameworks remains largely unexplored. The proposed method uses MC dropout as a regularization technique in prototypical networks in conjunction with the estimation of uncertainty. This approach addresses uncertainty estimation, which is an essential limitation in existing meta-learning algorithms. The main objective was to integrate prototypical networks with MC dropout by combining the efficiency of metric-based meta-learning with principled uncertainty estimation, offering a more robust and transparent learning environment to address overfitting in metric-based meta-learning frameworks.

II. METHODOLOGY

A. *Meta Learning*

Meta-learning originated from the aspiration to achieve general artificial intelligence, emphasizing the importance of rapid adaptation and learning in diverse and dynamic environments. The tasks in the meta-learning are divided into the support-set and query-set in the training phase. A random knowledge-based initialization θ sets the meta-model parameters before sampling several tasks from their distribution while using support example sets to develop task-specific parameters (ϕ) through the meta-model parameters. The meta-model parameters θ are updated based on the gradient of meta-loss with respect to generalization [17]:

$$L_{meta} = compute\_meta\_loss(\{L_i\}_{i=1}^{b}) \qquad (1)$$

B. *Prototypical Networks*

Prototypical networks [18] represent an innovative approach to few-shot learning that emerged from the convergence of metric learning and meta-learning paradigms. The fundamental concept lies in transforming input data into a semantically meaningful embedding space, where class representations are reduced to representative prototypes (Figure 1). The Mahalanobis distance is a widely used method in this framework through a defined matrix M. The distance formula determines the separation between points x_i and x_j as:

$$D_M(x_i, x_j) = \sqrt{(x_i - x_j)^T} \times M(x_i - x_j) \qquad (2)$$

Let the embedding function be represented as follows:

$$f_\varphi(x): R^d \to R^m \qquad (3)$$

where x is the input data point, φ denotes the learnable parameters of the embedding function, d is the dimensionality of the input space, and m is the dimensionality of the embedding space.

For a given class c, the prototype P_c is the mean vector of the embedded support set examples :

$$P_c = \frac{1}{|S_c|} \sum_{x_i \in S_c} f_\varphi(x_i) \qquad (4)$$

where P_c is the prototype for class c, S_c is the support set for class c, and $f_\phi(x_i)$ is the embedding of the support image x_i. These prototypes represent each class in the embedding space.

For a query example x_q, its embedding $f_\phi(x_q)$ is compared to each prototype P_c using a distance metric. The query point is classified using the SoftMax over negative distances parameters, tuned during training to reduce the negative log-likelihood [19, 20].

$$P\left(y = c \mid x_q \right) = \frac{\exp\left(-d(f_\phi(x_q), p_c)\right)}{\sum_{c'} exp\left(-d(f_\phi(x_q), p_{c'})\right)} \tag{5}$$

$$\mathcal{L} = -\frac{1}{|Q|} \sum_{(x_q, y_q) \in Q} \log P\left(y_q \mid x_q \right) \tag{6}$$

Finally, for prediction y^*, the network assigns each query point to the class whose prototype is the closest.

$$y^* = argmin_k d(f_\theta(x)) \tag{7}$$

MC dropout serves as a technique for neural network uncertainty and prediction estimation by interpreting dropout applications as part of Bayesian statistics. Each T stochastic forward pass applies a new dropout mask to the trained model. The predicted outcomes are combined to provide the average prediction, which represents the model's expected output. The variance of these predictions is then calculated to quantify predictive uncertainty, where a higher variance indicates greater uncertainty.

$$\hat{y}_t = f(x; W \odot Mt), \text{ for } t = 1, 2, \dots, T \tag{8}$$

C. Prototypical Networks with MC Dropout

The proposed method combines prototypical networks with MC dropout to deal with two key challenges in few-shot learning, i.e., quick learning from limited data and precise uncertainty estimation. In few-shot image classification, the task consists of a Support set (S) with labeled examples and a Query set (Q) with unlabeled examples. The method aims to categorize query samples based on their closeness to class prototypes, which serve as typical feature embeddings. The feature extractor $f(\theta)$ maps input images into a lower-dimensional embedding space.

$$\overline{z_j} = \sum_{t=1}^{T} z_j^t \tag{9}$$

The final query representation is calculated as the average of these stochastic embeddings. This method computes the squared Euclidean distance between the query embedding and each class prototype. MC dropout postulates for several probability estimates over t forward passes as opposed to a single deterministic estimate.

$$P\left(y_j = c \middle| x_j\right) = \frac{exp\left(-d_{j,c}\right)}{\sum_{c'} exp\left(-d_{j,c'}\right)} \tag{10}$$

The mean of these estimates is the final predicted probability for each class, and the variance across stochastic passes is used to assess uncertainty.

$$\hat{P}\left(y_j = c\right) \frac{1}{T} \sum_{t=1}^{T} P\left(y_j = c \middle| x_j\right)^t \tag{11}$$

Finally, the projected class, along with the uncertainty measure, gives important information about the model's confidence, enabling better decision-making.

$$\sigma^2\left(y_j\right) = \frac{1}{T} \sum_{t=1}^{T} \left(P\left(y_j = c \middle| x_j\right)^t - \hat{P}\left(y_j = c\right) \right)^2 \tag{12}$$

The predicted class for each query sample is defined as:

$$\widehat{y_j} = \arg\max_c \hat{P}\left(y_j = c\right) \tag{13}$$

Uncertainty estimation in classification is calculated using entropy through MC dropout, as shown in Figure 1. Entropy is calculated to assess the overall final uncertainty of the model prediction, with higher entropy signifying greater uncertainty.

$$H = -\sum_c p_c \log p_c \tag{14}$$

Algorithm 1 describes the overall method.

```
Algorithm 1: Prototypical Networks with MC
  dropout for few-shot image
  classification
Input: Support set $S$, Query set $Q$,
  Feature extractor $f_\theta(x)$ with MC dropout,
  Number of forward passes $T$,
  Set of Classes $C$
Output: Predicted labels $\widehat{y_j}$,
  Uncertainty estimates $\sigma^2(y_j)$

Procedure ComputePrototypes(S, T):
  for each class $c \in C$ do
    Compute prototype $p_c = \frac{1}{|S_c|}\sum_{x_i \in S_c} f_\phi(x_i)$
  end for
  return $p_c \ \forall c$
end procedure

Procedure PredictWithUncertainty($Q, p_c, T$):
  for each query sample $x_j \in Q$ do
    for $t = 1$ to $T$ do
      Apply dropout and compute embedding:
      $z_j^t = f_\theta(x_j)^t$
      Compute distances to prototypes
      Compute probabilities: $P(y_j = c \mid x_j)^t$
      using (10)
    end for
    Average Embeddings: $\overline{z_j} = \sum_{t=1}^{T} z_j^t$
    Compute mean prediction:
     $\hat{P}(y_j = c)\frac{1}{T}\sum_{t=1}^{T} P(y_j = c|x_j)^t$
    Predict class: $\hat{y}_j = \arg\max_c \hat{P}(y_j = c)$
    Compute uncertainty $\sigma^2(y_j)$ using (12)
    and Entropy $H$
  end for
  return $\{\hat{y}_j, \sigma^2(\hat{y}_j), H_j\}_{j \in Q}$
  end procedure
```

Unlike traditional prototypical networks that produce deterministic class assignments, this uncertainty-aware framework provides probabilistic inferences with confidence estimates. The primary difference stems from the stochastic embedding process. While conventional approaches compute a single embedding the proposed method generates diverse embeddings through dropout sampling. The mean averaging operation yields more robust representations than single-pass embeddings, while the variance and entropy enhance prediction uncertainty. This enhancement proves advantageous in few-shot scenarios where limited support samples make models susceptible to overconfident predictions on uncertain queries.

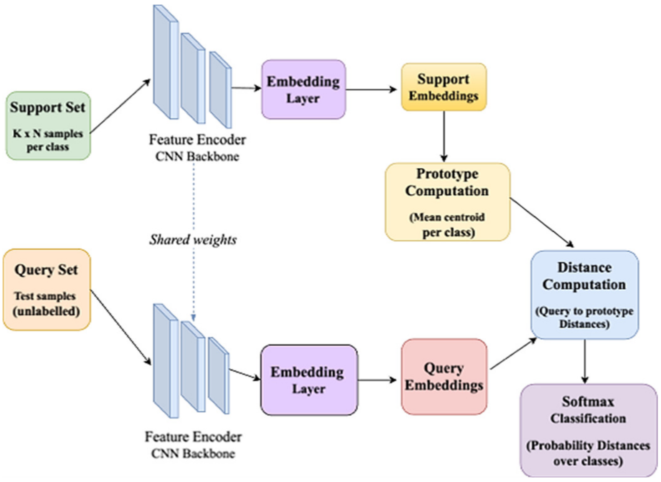

Fig. 1. Prototype-based few-shot learning, where query images are classified by comparing embeddings to the class prototypes.

III. RESULTS

A. *Datasets*

The proposed model was evaluated on four large datasets to assess performance in diverse domains and data distribution scenarios. The review of initial visual performance utilized the MNIST [21] benchmark dataset. An extensive experimental evaluation of few-shot learning performance involved the Mini-ImageNet [22] dataset. The CIFAR-10 [23] dataset was used as a test control to analyze how the model processed complex images with different ranges of variability. The TrashNet [24] dataset is a collection of waste images, which served to evaluate the practical capabilities of the model in real-world scenarios.

B. *Experimental Results*

The evaluation was conducted in 5-way 1-shot and 5-way 5-shot scenarios, where the model must classify new classes with a limited number of samples in each respective class. Five-way five-shot accuracy is an imperative metric in evaluating meta-learning algorithms in few-shot learning, where the aim is to classify images into one of five categories with only five examples per class during training the model. The results show that MC dropout improves generalization accuracy, implying that adding uncertainty estimation benefits the ability of prototypical networks to adapt to novel tasks. From the accuracy comparisons, it can be observed that MC dropout consistently leads to higher accuracy across datasets. The highest accuracy across all scenarios was achieved on MNIST, which aligns with its straightforward visual patterns. In TrashNet, which includes a higher level of real-world noise, significant gains were achieved with MC dropout, reinforcing the idea that uncertainty-aware meta-learning is advantageous for high-variance or ambiguous data distributions.

TABLE I. COMPARISON OF 5-WAY 1-SHOT AND 5-SHOT CLASSIFICATION ACROSS DATASETS, WITH AND WITHOUT MC DROPOUT

	Accuracy			
Dataset	**1-Shot (No dropout)**	**1-shot (MC dropout)**	**5-shot (No dropout)**	**5-shot (MC dropout)**
MNIST	85.12	88.35	84.13	92.06
CIFAR-100	64.31	65.76	69.08	73.64
Mini ImageNet	49.51	52.24	46.10	58.56
TrashNet	70.17	71.33	76.43	82.13

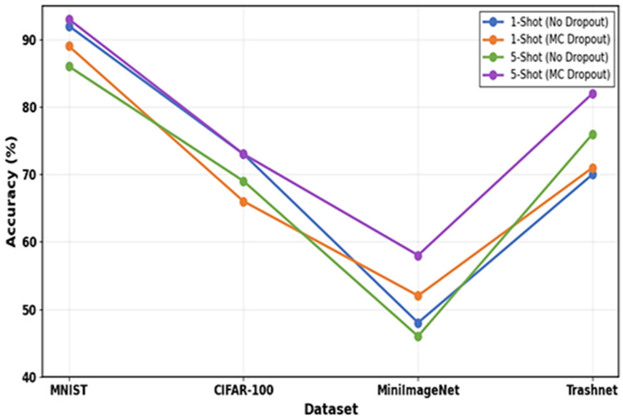

Fig. 2. Generalization accuracy comparison of 5-way 1-shot and 5-shot Image classification with and without MC dropout.

C. *Cross-Domain Few-Shot Classification on TrashNet Dataset*

To evaluate the effectiveness of the proposed method with respect to enhancing the generalization of metric-based meta-learning models, a cross-domain experiment was conducted using an unfamiliar meta-testing set. The cross-domain scenario involves meta-training on the Mini-ImageNet dataset and meta-testing on the TrashNet dataset to assess the model's ability to generalize to unknown domains. The training process focuses on refining the model parameters on the Mini-ImageNet dataset, followed by a refinement phase on the TrashNet dataset for waste classification. Despite having very few annotated samples, the algorithm improved both generalization accuracy and efficiency.

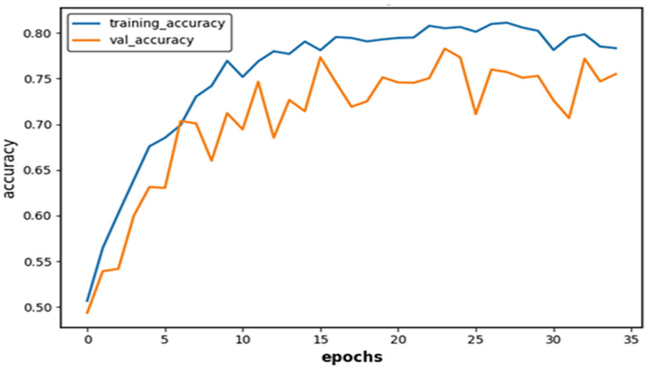

Fig. 3. Accuracy of the proposed algorithm using prototypical networks with the proposed MC dropout algorithm.

Figure 4 illustrates a constant decline in training loss, suggesting effective learning, indicating that the model can generalize meticulously to new data. Mini-ImageNet allows the model to learn the representations that contribute to simplified initialization for future tasks. After the meta-training phase, the TrashNet dataset was employed to fine-tune and evaluate the performance of the model for waste classification based on the respective waste type. A lower entropy value signifies that the model is confident in its decision, whereas a higher entropy value indicates uncertainty in its decision. The violin shape in Figure 5 illustrates exactly how the entropy values are distributed within each category. For samples that were correctly classified, the plot widens at lower entropy levels, indicating that most of these predictions were made with high confidence. In contrast, the distribution of misclassified data shifts toward higher values. The cross-domain findings substantiate real-world applicability for waste management applications. Through meta-training on real images and fine-tuning with minimal waste samples, the framework facilitates expedited implementation in unseen query tasks of the meta testing phase without the need for exhaustive data acquisition. The uncertainty measures facilitate the system to flag questionable samples for further human verification while maintaining classification trustworthiness for real-world scenarios where there are chances of imbalance in the dataset.

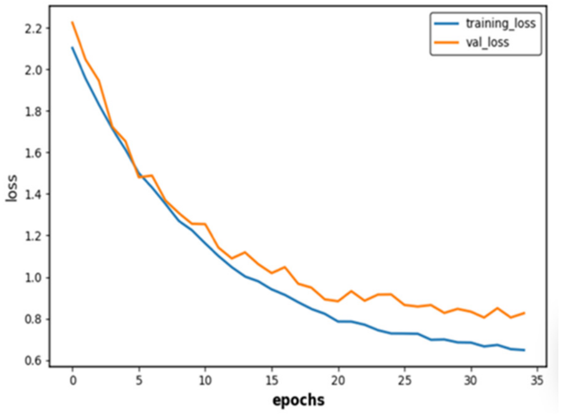

Fig. 4. Validation and training loss using prototypical networks with the proposed MC dropout algorithm.

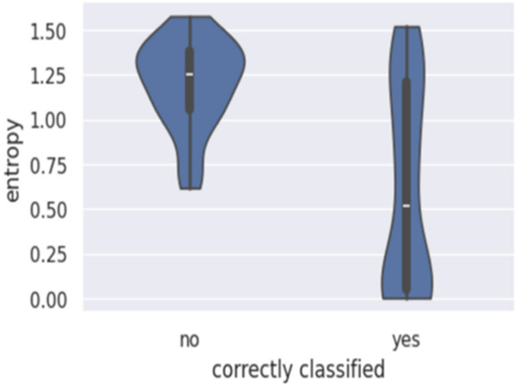

Fig. 5. Violin plots illustrating the entropy distribution for correctly and incorrectly classified samples.

IV. CONCLUSION

Existing metric-based meta-learning algorithms exhibit excellence in quick adaptation to new tasks, but also result in overconfident predictions that constrain their implementation in high-stakes domains due to a lack of uncertainty quantification. Compared to existing meta-learning approaches, the proposed method demonstrates competitive performance, offering uncertainty quantification competencies that are absent in standard meta learning algorithms, such as prototypical networks, matching networks, and Model-Agnostic Meta-Learning (MAML). Hence, combining the dropout results, the proposed method achieved 58.56% accuracy in a 5-way 5-shot scenario on the Mini-ImageNet dataset after integrating MC dropout, compared to 46.10% with baseline prototypical networks (without MC dropout). To contextualize these results within the broader meta-learning domain, the MAML algorithm reported approximately 48.7% for 5-way 1-shot on Mini-ImageNet, while matching networks achieved approximately 43.6%. Recent transformer-based approaches, such as ViT-ProtoNet, have also shown enhanced performance on Mini-ImageNet benchmarks, but these methods require significantly more computational resources and do not offer uncertainty quantification. The model also shows relative consistency on other datasets, scoring 92.06% 73.64%, and 82.13% on 5-shot scenarios on MNIST, CIFAR-100, and TrashNet, compared to 84.13%, 69.08%, and 76.43% for the baseline. The model retains proportional accuracy, maintaining entropy-based measures of uncertainty estimation for each prediction, which facilitates estimation of low-confidence samples that need human verification.

These results highlight the importance of meta-learning and its varied practical applications. The developed ability to combine fast adaptation and reliable uncertainty estimation fills a significant gap in existing meta-learning methods. This advancement paves the way for secure applications of few-shot learning in image classification scenarios, especially medical diagnostics, autonomous vehicles, and environmental monitoring, where accurate predictions are essential. Future research directions entail extending this framework to include multi-modal learning scenarios and evaluating the performance of other meta-learning algorithms in important fields, such as industrial systems, fault detection, and real-time anomaly detection in cybersecurity environments, where the consequences of misclassification are especially high.

ACKNOWLEDGMENT

The authors acknowledge the Department of Information Science and Engineering, Ramaiah Institute of Technology, Bengaluru, for providing the computational resources and infrastructure necessary to carry out this research.

REFERENCES

[1] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a Few Examples: A Survey on Few-shot Learning," *ACM Computing Surveys*, vol. 53, no. 3, Mar. 2020, Art. no. 63, https://doi.org/10.1145/3386252.

[2] A. Bansal, R. Sharma, and M. Kathuria, "A Systematic Review on Data Scarcity Problem in Deep Learning: Solution and Applications," *ACM Computing Surveys*, vol. 54, no. 10, Jan. 2022, Art. no. 208, https://doi.org/10.1145/3502287.

[3] T. M. Hospedales, A. Antoniou, P. Micaelli, and A. J. Storkey, "Meta-Learning in Neural Networks: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 5149-5169, 2021, https://doi.org/10.1109/TPAMI.2021.3079209.

[4] H. Gharoun, F. Momenifar, F. Chen, and A. H. Gandomi, "Meta-learning Approaches for Few-Shot Learning: A Survey of Recent Advances," *ACM Computing Surveys*, vol. 56, no. 12, Apr. 2024, Art. no. 294, https://doi.org/10.1145/3659943.

[5] A. Vettoruzzo, M. R. Bouguelia, J. Vanschoren, T. Rögnvaldsson, and K. Santosh, "Advances and Challenges in Meta-Learning: A Technical Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 7, pp. 4763–4779, July 2024, https://doi.org/10.1109/TPAMI.2024.3357847.

[6] J. Wang *et al.*, "Neuromodulated Meta-Learning." arXiv, Nov. 11, 2024, https://doi.org/10.48550/arXiv.2411.06746.

[7] M. Abdar *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, Dec. 2021, https://doi.org/10.1016/j.inffus.2021.05.008.

[8] A. G. Khoee, Y. Yu, and R. Feldt, "Domain generalization through meta-learning: a survey," *Artificial Intelligence Review*, vol. 57, no. 10, Sept. 2024, Art. no. 285, https://doi.org/10.1007/s10462-024-10922-z.

[9] W. Guo, F. Zhuang, X. Zhang, Y. Tong, and J. Dong, "A comprehensive survey of federated transfer learning: challenges, methods and applications," *Frontiers of Computer Science*, vol. 18, no. 6, Dec. 2024, Art. no. 186356, https://doi.org/10.1007/s11704-024-40065-x.

[10] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," in *Proceedings of The 33rd International Conference on Machine Learning*, June 2016, New York, NY, USA, pp. 1050–1059.

[11] A. Lemay *et al.*, "Improving the repeatability of deep learning models with Monte Carlo dropout," *npj Digital Medicine*, vol. 5, no. 1, Nov. 2022, Art. no. 174, https://doi.org/10.1038/s41746-022-00709-3.

[12] J. Xie *et al.*, "Advanced Dropout: A Model-Free Methodology for Bayesian Dropout Optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 4605–4625, Sept. 2022, https://doi.org/10.1109/TPAMI.2021.3083089.

[13] A. Shamsi, H. Asgharnezhad, A. Tajally, S. Nahavandi, and H. Leung, "An Uncertainty-aware Loss Function for Training Neural Networks with Calibrated Predictions." arXiv, Feb. 06, 2023, https://doi.org/10.48550/arXiv.2110.03260.

[14] X. Liang *et al.*, "R-Drop: Regularized Dropout for Neural Networks," in *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 10890–10905.

[15] M. Hasan, I. Hossain, A. Rahman, and S. Nahavandi, "Controlled Dropout for Uncertainty Estimation," in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, July 2023, HI, USA, pp. 973–980, https://doi.org/10.1109/SMC53992.2023.10394101.

[16] M. Mujiyanto, A. Setyanto, K. Kusrini, and E. Utami, "Swin Transformer with Enhanced Dropout and Layer-wise Unfreezing for Facial Expression Recognition in Mental Health Detection," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 19016–19023, Dec. 2024, https://doi.org/10.48084/etasr.9139.

[17] C. Finn, P. Abbeel, and S. Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks," in *Proceedings of the 34th International Conference on Machine Learning*, July 2017, Sydney, Australia, pp. 1126–1135.

[18] J. Snell, K. Swersky, and R. Zemel, "Prototypical Networks for Few-shot Learning," in *Advances in Neural Information Processing Systems*, 2017, vol. 30, pp. 4077–4087.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[20] X. Li, X. Yang, Z. Ma, and J. H. Xue, "Deep metric learning for few-shot image classification: A Review of recent developments," *Pattern Recognition*, vol. 138, June 2023, Art. no. 109381, https://doi.org/10.1016/j.patcog.2023.109381.

[21] L. Deng, "The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, Aug. 2012, https://doi.org/10.1109/MSP.2012.2211477.

[22] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching Networks for One Shot Learning," in *Advances in Neural Information Processing Systems*, 2016, vol. 29.

[23] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," University of Toronto, Canada, Technical Report, 2009.

[24] M. Yang and G. Thung, "Classification of Trash for Recyclability Status," Stanford University, Course Project CS 229.