

A Systematic Literature Review of Meta-Learning Models for Classification Tasks

Jackson Kamiri
Murang'a University of
Technology, Kenya

Geoffrey Mraiga
Murang'a University of
Technology, Kenya

Aaron Oirere
Murang'a University of
Technology, Kenya

Abstract: Meta-learning is a field of learning that aims at addressing the challenges of conventional machine learning approaches such as learning from scratch for every new task. The main aim of this study was to do a systematic literature review of the existing meta-learning models that have been developed, published, and can be used for classification tasks. Systematic literature review method was used, employing a search of journal articles and publications of conference proceedings. The process involved data collection, analysis, and reporting of the results. To achieve the objective, 30 primary papers published since 2016 and relevant to classification tasks in meta-learning were considered. Data was extracted from the papers, then the following was analyzed in each model as presented in the papers; techniques used, the contribution, and the research gap. Although a lot has been done so far in Meta-learning, the existing models are not yet optimal. They still have challenges in few-shot learning, computation time complexity, difficulty in continual learning, and generalizability across multiple related tasks during transfer learning.

Keywords: Machine Learning; Meta-Learning; Few-Shot Learning; Transfer-Learning.

1.0 INTRODUCTION

Meta-learning also known as learning-to learn is a field of learning that aims at addressing the challenges of conventional machine learning approaches such as generalization and learning from scratch for every new task [1]. Through meta-learning, deep learning models are able to learn from a variety of related tasks and then use transfer learning to enable them to solve a new but related task with just a few data samples also known as few-shot learning. This enables machine learning developers to develop models that are robust and with high-level performance even in areas where labeled training data is limited such as in the medical field [2]. Meta-learning can be used to perform a variety of tasks such as regression, classification, and reinforcement learning. However, this study mainly focuses on classification tasks using Meta-learning algorithms.

Finn et al. [1], further argues that meta-learning enables machines to gain state-of the art learning capabilities almost similar to that of a natural human being. The main aim of this study is to do a systematic literature review of the existing meta-learning models that have been developed, published, and can be used for classification tasks. According to Kitchenham [3] a systematic literature review is means through which available research relevant to a particular research question, phenomenon or topic of interest is evaluated and interpreted. Unlike ordinary literature review, Systematic literature review is evidence-based in the sense that its approach is well documented, scientific and reproducible.

Previous studies such [4] and [5] have done a survey on Meta-learning. However, we fault the above studies since they did not do a systematic literature review. Our contribution therefore, is that we have used a scientific

approach which is evidence-based and guided by the guidelines of Kitchenham [3].

2.0 METHODOLOGY

We employed systematic literature review approach in reviewing meta-learning models for classification tasks.

2.1 Research Questions

This study was guided by three research questions (RQ) which are:

1. Which techniques have the existing meta-learning models used?
2. What problems have the models addressed?
3. What are the gaps that the existing models have left that can be addressed by future studies?

2.2 Inclusion criteria

For a paper to be included in the review it needs to meet the following criteria; first, published between 2016 to 2021. Second, published in English language. Third, must be a peer-reviewed journal paper or published conference proceedings. Fourth, the paper is available in google scholar. Fifth, must be using meta-learning to perform classification tasks. Sixth, has a well-documented methodology, contribution and results. Seventh, the paper must be cited in google scholar citations.

2.3 Exclusion Criteria

The papers that did not meet the criteria discussed in section 3.2 were excluded from the study. On top of the above-discussed criteria, in cases where more than one version of the same paper were available, the most comprehensive version was included while the others were excluded.

2.4 Identification of papers

To identify potential papers for this study, we combined three techniques which are; searching for papers in google scholar search engine, identification of papers using references from included studies, and manual search in IEEE

Explore repository. The key words combination that we used in the search include: “Meta-learning” “Meta-learning for classification”, “classification using meta-learning “, “Transfer learning”, and “Learning-to-learn”. The first keyword provided the largest output of papers in google scholar. We used several search key words to overcome the threat of omission where the initial key word was not very conspicuous in a study. The total number of papers obtained from the search was 58 papers.

2.5 Quality Assessment.

According to Kitchenham [3] quality is sometimes subjective. Therefore, we set our quality threshold to be the capability of a paper to answer all our research questions. To determine how well each paper answered our research questions, we used the following criteria:

First, determining if the study is a meta-learning classification task: in this we read the abstract and introduction sections of the paper to know if the paper was performing classification using meta-learning techniques.

Second, establishing the contribution of the paper: we scrutinized the papers thoroughly to determine the problem they have solved and how they have contributed to the field of meta-learning. In this we considered any improvement to the studies that previously existed as noted in the literature survey of the paper.

Third, determining if the paper has documented its methodology properly: we ensured that all papers included have a clear methodology that documents the following critical elements; the techniques used, the dataset used, the results obtained, and discussion of the results.

After subjecting all the papers obtained to the inclusion and quality assessment criteria, only 30 of the 58 papers met the needed threshold. Therefore, the results that we present in this paper are based on the 30 analyzed papers. Table 1 shows a summary of the number of papers sourced and how the numbers fared in each stage. The table follows the structure proposed by [6]

Table 1: Papers Included after Applying Quality Criteria

Total Sourced	Numbers of papers that failed Stage 1	Numbers of papers that failed Stage 2	Numbers of papers that failed Stage 3	Number of papers that qualified all stages
58	20	5	3	30

2.6 Data extraction and analysis

Data extraction was guided by the research questions. We developed data extraction forms which we used to extract data from each primary study. Attached in appendix 2 is a sample data extraction form. The researchers concentrated with what the papers reported rather than their personal interpretation of what the papers documented. The key elements that were extracted from each paper include:

RQ1, the techniques used in the paper: we focused on the specifics of the algorithms, techniques, datasets, and the general research approach used in the paper. For instance, [7] combined gradient-based meta-learning with model-based meta-learning. We also considered the application area in which the paper was applied, this was mainly informed by the dataset used.

RQ2, Contribution of the paper: we considered the value that each paper brings on-board. In this we captured what the authors of each paper have documented as their contribution. In cases where comparison was available, we considered contribution in comparison with previous studies in the same context such as in the case of [8] which was compared to [1]. All the papers reviewed had a contribution which is discussed in the results sections. We also extracted quantitative data in the form of performance metrics of each paper with respect to the experiment conducted and the datasets used.

RQ3, limitations of each paper: here we considered the gap that each paper has left out. Then we checked if any other paper has filled the gap by solving the limitations identified. We then documented in our results the gap that has not yet be covered by either of the papers analyzed.

2.7 Deviations from protocol

A review protocol defines the methods that will be used to undertake a specific systematic literature review [6]. The protocol is composed of all the elements of the review and some additional planning information. This study followed the protocol as defined in the methodology section of the paper. The researchers committed to follow the protocol in order to ensure that the results of the study are not influenced by researchers' bias.

3.0 RESULTS

Appendix 1: shows the papers that met the inclusion, exclusion, and quality assessment criteria. This section demonstrates how the researchers answered each of the research questions that formed the basis of our study.

3.1 RQ1: Which techniques have the existing meta-learning models used

The results of this study demonstrate that each of the 30 studies reviewed used techniques which incorporated a baseline and a technique unique to the specific problem that the paper intended to solve. For instance Goldblum et al. (2020) used adversarial querying together with the baseline techniques to achieve the objective of the model while [10] used hierarchical structuring on top of the base-structure. The baseline across all the papers reviewed involved the following techniques Neural network, meta-learner, multi-task learning, an optimizer for adaptation, and few-shot learning.

Another notable theme is that the model Agnostic Meta-Learning proposed by [1] has formed a standard that 86% of the analyzed papers extended. Of these 86% of the papers, 1 paper by [2] extended an extension of MAML known as

Reptile proposed by [11]. Therefore, these models that extended MAML used the MAML architecture as their baseline and then added new techniques to it. For instance [12] introduced noise gradient-based learning as way of extending MAML while [7] extended MAML by combining model-based learners with gradient-based learning.

All the papers considered in this research used secondary datasets to develop the models. 73.4% used more than one dataset while 26.6% used only one dataset.

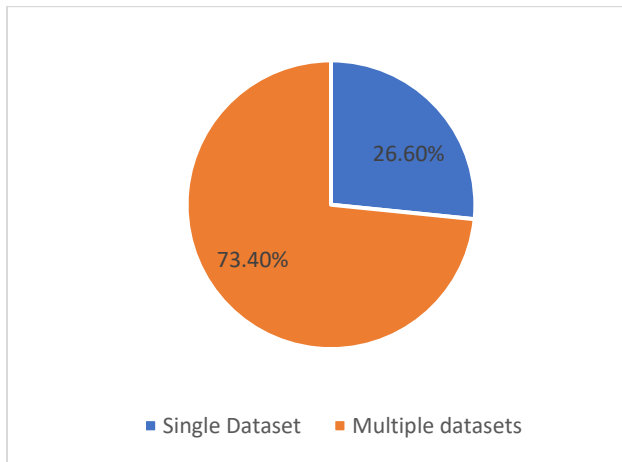


Figure 1: Data set Usage analysis

The most used datasets by the reviewed papers are Omniglot (36.67%), MiniImagnet (56.67%), MINST (16.67%), and Cifar-10+Cifar-100(16.67%).

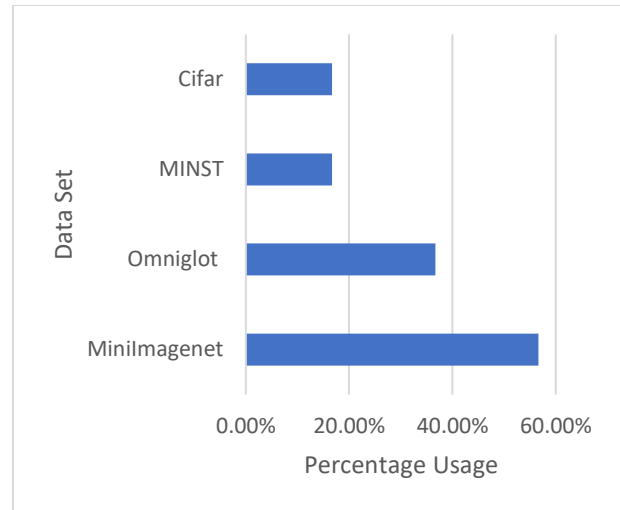


Figure 2: Most used datasets

The other datasets used are unique to the papers that used them. Papers such as [2] used the MiniImagnet dataset to pretrain the model then adapted the model through transfer learning to fit the Diabetic Retinopathy dataset. Also [13] used a pretrained ResNet50 which was trained on ImageNet then finetuned through transfer learning to fit labeled Retina Images.

3.2 RQ2: Which Problem have the papers solved.

Each of the papers analyzed in this section made some considerable contribution to the field of meta-learning in classification tasks. Some papers improved on classification accuracy while others introduced new and better approaches towards solving classification problems in meta-learning. In order to present the contributions of each paper clearly and in a summarized way, we have used paper codes to refer to the specific papers in appendix 1.

S01: Developed an optimization by gradient descent through learning rather than hand-crafted optimizers. S02: combined gradient-based learning with external memory modules. S03: proposed an LSTM based meta-learner model to learn the exact optimization algorithm used to train another learner neural network classifier in the few-shot regime. S04: proposed a model agnostic meta-learning model. S05: proposed a model that is able to acquire meta-learning prior for new tasks for multimodal task distribution. S06: proposed a model that solved catastrophic forgetting in

meta-learning. S07: proposed a model that efficiently obtains a task posterior of a novel task. S08: proposed a model that improved MAML by introducing task robustness.

S09: solved the problem of task ambiguity in few-shot learning. S10: proposed an online hyperparameter adaptation scheme that eliminates the need to tune learning rates and meta-learning hyperparameters of the MAML. S11: proposed a model that dynamically updates the learning rate. S12: introduced a model that decouples meta-gradient computation from the choice of inner loop optimizer. S13: incorporated latent embedding optimization in gradient-based learning. S14: proposed a model in which they designed meta-regularization objective using information theory. S15: proposed an algorithm that hierarchically structures the transferrable knowledge into different clusters of tasks. S16: proposed an algorithm that is capable of maintaining an equilibrium between all the encountered tasks.

S17: proposed a model that scales without overfitting and is robust to task specific learning rate. S18: Combined hierarchical Bayesian models and gradient-based models. S19: improved Reptile by training it on transfer learning. S20: introduced unsupervised meta-learning to MAML. S21: combined meta-learning with the traditional fine-grained classification algorithms to improve on classification. S22: identified hyperparameters that generate optimal performance of MAML in image classification. S23: Expanded on MAML by demonstrating that first-order meta-learning algorithms perform well on some well-established benchmarks for few shot image classifications and provided theoretical analysis aimed at understanding why those algorithms worked.

S24: Proposed a model that uses trained linear classifiers as base learners to learn representations for few-shot learning. S25: advanced few-shot classification paradigm towards a scenario where unlabeled examples are also available within each episode of learning. S26: proposed adversarial querying algorithm that is robust to adversarial examples and perform well in few-shot learning. S27: proposed a model to guide the learning of network parameters so that they are optimal

for adapting to the target fine-grained classification task. S28: proposed a model that stores local states that extract the experience information from the seen task in transfer learning. S29: Proposed a meta-learning model that uses soft noisy labels. The model is trained using conventional approach and transfer learning. S30: the model fully integrates neural architecture search with gradient-based meta-learning.

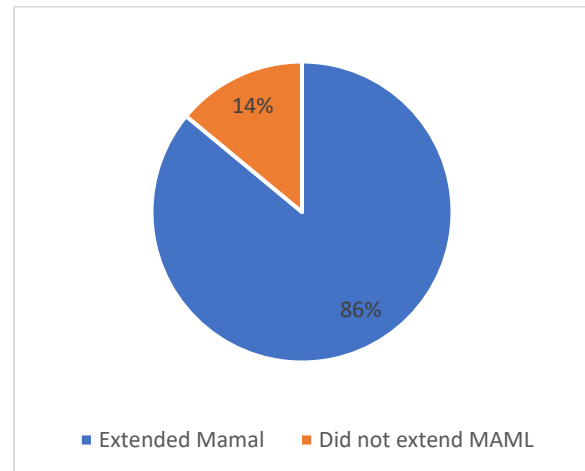


Figure 3: Analysis of papers that solved a problem that involved extending MAML

3.3 RQ3 What are the gaps that the existing models have left that can be addressed by future studies?

To answer this question, we considered the gaps or limitations of each of the 30 papers reviewed. Apart from our detailed analysis, we as well considered future works proposed by the authors of the papers to be part of the gap. We also considered whether or not either of the papers in the pool solved gaps left by other papers in the pool. For instance, the proposed future work by S04 was addressed by S05. This section therefore, reports outstanding gaps that have not yet been solved by either of the papers considered in the study.

First, S07 uses Stein variational gradient descent (SVGD) and hierarchical Bayesian and thus it does not have the full coverage of the task-posterior, a neural network would deliver better in this. Second, still in S07 the model suffers the shortcomings of ensemble approaches such as space/time complexity proportional to the number of particles. Third,

SVGD in S07 is limited in that its performance is sensitive to the parameters of the kernel function[14]. Fourth the model developed by S09 provides impoverished estimator of posterior variance, thus, its effectiveness in gauging where task have different degrees of uncertainty is low[12]. Fifth, in S11 the method has not yet achieved convergence results of non-transitioning variant of hyper gradient descent, and also has not established the convergence rate [15].

Sixth, S13 the researchers proposed a future work that replaces the pre-trained feature extractor with one learned jointly through meta-learning, or using LEO for tasks in reinforcement learning or with sequential data [16]. Seven, in S18 the researcher records that the Laplace approximation is inaccurate in cases where the integral is highly skewed, or is not unimodal and thus is not amenable to approximation by a single Gaussian mode [17].

Eight, S19 the researcher proposes that in future an advanced neural network architecture can be used for transfer learning[2]. Also, the researcher uses default learning rates rather than dynamic learning rates thus limiting the capability of the model to achieve optimal performance quickly. Nine S20 the model traded off accuracy and reduction in the number of labeled images required. Therefore, it cannot be useful where classification accuracy is essential [18].

Ten, S24 proposed that future work can explore other convex base-learners such as kernel Support Vector Machines [19]. Eleven, S23 the researchers record that further research should: pay close attention to transduction's use of batch normalization during testing, seek to understand to what extent SGD automatically optimizes for generalization, explore if regularization can improve few-shot learning[11]. Twelve, S25 the researchers proposed that future work can extend their work by incorporating fast weights[20].

Thirteen S29 the model noisy labels impairs the convergence of SGD. Also, the model only does binary classification yet Diabetic retinopathy is multi-class classification task[13]. Fourteen, all studies that have used gradient-based learners have faced a generalization challenge as a result of poor convergence of the loss function in both the local and global

minima[1], [21]. Thirteen models such as S18 consumed a lot of execution time [17].

3.1 Discussion

This research demonstrates that meta-learning has become a better alternative to deep learning in performing classification tasks. This is mainly due to the capabilities of meta-learning to leverage on previous knowledge during transfer learning, perform automatic hyperparameter optimization and learn from few data samples. According to Fin et al [1], the capability to leverage on previous knowledge is the hallmark of human intelligence and thus meta-learning is a clear path towards artificial general intelligence.

The study has also demonstrated that the analyzed models use baseline which mainly constitutes of a neural network, meta-learner, multi-task learning, and an optimizer. Then, researchers build their models on the basis of this baseline. Meta-learning will continue to attract a lot of research interest among researchers since there are many grey areas as enumerated in section 3.3 that are yet to be addressed.

3.2 Limitations of the study

This study only covered Meta-learning models used for classification tasks. Therefore, it has not covered other areas in which meta-learning has been applied such as in reinforcement learning, regression, speech recognition, and text analytics.

4.0 CONCLUSION

In this study we have conducted a thorough systematic literature review of meta-learning models used for classification task. This work is part of continuing research. This study forms a strong foundation for researchers interested in the field of Meta-Learning since it provides sufficient information on the current status of meta-learning research in classification tasks. The study has demonstrated clearly how the analyzed models fair in the research questions of this study.

In future researchers can consider solving the gaps that have been identified in section 3.3 of this paper.

References

- [1] C. Finn, P. Abbeel, and S. Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” *34th Int. Conf. Mach. Learn. ICML 2017*, vol. 3, pp. 1856–1868, 2017.
- [2] M. Welling, “Meta-Learning for Medical Image Classification,” no. Midl, pp. 7–9, 2018.
- [3] B. Kitchenham, “Kitchenham, B. : Guidelines for performing Systematic Literature Reviews in software engineering . EBSE Technical Report EBSE-2007-01 Guidelines for performing Systematic Literature Reviews in Software Engineering,” no. January 2007, 2021.
- [4] T. M. Hospedales, A. Antoniou, P. Micaelli, and A. J. Storkey, “Meta-Learning in Neural Networks: A Survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–20, 2021, doi: 10.1109/TPAMI.2021.3079209.
- [5] I. Khan, X. Zhang, M. Rehman, and R. Ali, “A Literature Survey and Empirical Study of Meta-Learning for Classifier Selection,” *IEEE Access*, vol. 8, pp. 10262–10281, 2020, doi: 10.1109/ACCESS.2020.2964726.
- [6] T. Hall, S. Beecham, D. Bowes, D. Gray, and S. Counsell, “A systematic literature review on fault prediction performance in software engineering,” *IEEE Trans. Softw. Eng.*, vol. 38, no. 6, pp. 1276–1304, 2012, doi: 10.1109/TSE.2011.103.
- [7] R. Vuorio, S. H. Sun, H. Hu, and J. J. Lim, “Multimodal model-agnostic meta-learning via task-aware modulation,” *Adv. Neural Inf. Process. Syst.*, vol. 32, no. NeurIPS, pp. 1–22, 2019.
- [8] L. Collins, A. Mokhtari, and S. Shakkottai, “Task-Robust Model-Agnostic Meta-Learning,” pp. 1–30, 2020.
- [9] M. Goldblum, L. Fowl, and T. Goldstein, “Adversarially robust few-shot learning: A meta-learning approach,” in *Advances in Neural Information Processing Systems*, 2020, vol. 2020-Decem, no. 1, pp. 1–15.
- [10] A. A. Liu, Y. T. Su, W. Z. Nie, and M. Kankanhalli, “Hierarchical Clustering Multi-Task Learning for Joint Human Action Grouping and Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 102–114, 2017, doi: 10.1109/TPAMI.2016.2537337.
- [11] A. Nichol, J. Achiam, and J. Schulman, “On First-Order Meta-Learning Algorithms,” pp. 1–15.
- [12] C. Finn, K. Xu, and S. Levine, “Probabilistic model-agnostic meta-learning,” *Adv. Neural Inf. Process. Syst.*, vol. 2018-Decem, no. NeurIPS, pp. 9516–9527, 2018.
- [13] G. Algan, I. Ulusoy, Ş. Gönül, B. Turgut, and B. Bakbak, “Deep Learning from Small Amount of Medical Data with Noisy Labels: A Meta-Learning Approach,” 2020.
- [14] J. Yoon, T. Kim, O. Dia, S. Kim, Y. Bengio, and S. Ahn, “Bayesian Model-Agnostic Meta-Learning,” no. NeurIPS, pp. 1–11, 2018.
- [15] D. Mart, F. Wood, R. Cornish, and M. Schmidt, “ONLINE LEARNING RATE ADAPTATION WITH,” no. 2015, pp. 1–11, 2018.
- [16] A. A. Rusu *et al.*, “META-LEARNING WITH LATENT EMBEDDING OPTIMIZATION,” pp. 1–17, 2019.
- [17] E. Grant, C. Finn, S. Levine, T. Darrell, T. Griffiths, and C. Sciences, “Reinforcement Learning,” pp. 1–13, 2017.
- [18] S. Khodadadeh, L. Bölöni, and M. Shah, “Unsupervised meta-learning for few-shot image classification,” in *Advances in Neural Information Processing Systems*, 2019, vol. 32, no. NeurIPS.

- [19] K. Lee and S. Maji, “Meta-Learning with Differentiable Convex Optimization.”
- [20] M. Ren *et al.*, “Meta-learning for semi-supervised few-shot classification,” in *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018, pp. 1–15.
- [21] L. Zintgraf, K. Shiarlis, V. Kurin, K. Hofmann, and S. Whiteson, “Fast Context Adaptation via Meta-Learning,” no. 2018, 2019.
- [22] M. Andrychowicz *et al.*, “Learning to learn by gradient descent by gradient descent,” *Adv. Neural Inf. Process. Syst.*, no. Nips, pp. 3988–3996, 2016.
- [23] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, “Meta-Learning with Memory-Augmented Neural Networks,” *33rd Int. Conf. Mach. Learn. ICML 2016*, vol. 4, pp. 2740–2751, 2016.
- [24] S. Ravi and H. Larochelle, “Optimization as a model for few-shot learning,” *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, pp. 1–11, 2017.
- [25] R. Vuorio, D.-Y. Cho, D. Kim, and J. Kim, “Meta Continual Learning,” 2018.
- [26] H. S. Behl, A. G. Baydin, and P. H. S. Torr, “Alpha MAML: Adaptive Model-Agnostic Meta-Learning,” 2019.
- [27] A. Rajeswaran, S. M. Kakade, C. Finn, and S. Levine, “Meta-learning with implicit gradients,” *Adv. Neural Inf. Process. Syst.*, vol. 32, no. NeurIPS, pp. 1–12, 2019.
- [28] M. Yin, G. Tucker, M. Zhou, S. Levine, and C. Finn, “M -l m,” pp. 1–21, 2020.
- [29] H. Yao, Y. Wei, J. Huang, and Z. Li, “Hierarchically Structured Meta-learning,” 2019.
- [30] J. Rajasegaran, S. Khan, M. Hayat, F. S. Khan, and M. Shah, “iTAML : An Incremental Task-Agnostic Meta-learning Approach,” pp. 13588–13597.
- [31] X. Ruan, H. Liu, W. Pang, and S. Lu, “Fine-grained Classification Algorithm based on,” pp. 2019–2022, 2019.
- [32] Y. Zhang, H. Tang, and K. Jia, “Fine-Grained Visual Categorization using Meta-Learning Optimization with Sample Selection of Auxiliary Data.”
- [33] D. Alp, E. Acar, R. Zhu, and V. Saligrama, “Memory Efficient Online Meta Learning,” 2021.
- [34] T. Elsken, B. Staffler, J. H. Metzen, and F. Hutter, “Meta-Learning of Neural Architectures for Few-Shot Learning,” pp. 12365–12375.

Appendix 1

Table 2: List of papers Reviewed

Paper Code	Citation
S01	[22]
S02	[23]
S03	[24]
S04	[1]
S05	[7]
S06	[25]
S07	[14]
S08	[8]
S09	[12]
S10	[26]
S11	[15]
S12	[27]
S13	[16]
S14	[28]
S15	[29]
S16	[30]
S17	[21]
S18	[17]
S19	[2]
S20	[18]
S21	[31]
S22	[31]
S23	[11]
S24	[19]
S25	[20]
S26	[9]
S27	[32]
S28	[33]
S29	[13]
S30	[34]

Appendix 2:

Table 3: Sample Data Extraction Form

Paper Code	S06
Title	Meta Continual Learning
Authors	Risto Vuorio, Dong-Yeon Cho, Daejoong Kim, and Jiwon Kim
Techniques Used	ANN (predictor) which updated a loss function on current and previous tasks.
Dataset Used	MINST
Main Contribution	proposed a meta-learning model that aimed at solving catastrophic forgetting problem in deep learning
Gaps/ Limitations	Although this model achieved continual learning, the researchers record that it was faced with the challenge of learning to optimize. Poor optimization affected the generalizability of the model.