Tampere University

Md Enamul Haque

INVESTIGATION OF META-LEARNING TO ENHANCE SUPERVISED LEARNING

Master of Science
Information and Communication Technology
Statistical Data Analytics
Supervisor: Prof. Frank Emmert-Streib
May 2024

ABSTRACT

Md Enamul Haque: Investigation of meta-learning to enhance supervised learning
Master of Science
Tampere University
Statistical Data Analytics
May 2024

This thesis investigates the application of meta-learning techniques to enhance the supervised machine-learning model's adaptability, efficiency, and performance. Meta-learning, or "learning to learn," enables models to refine their learning processes based on past experiences, promising significant improvements over supervised machine-learning methods. The primary objectives of this research were to optimize supervised machine learning algorithms and enable these models to manage and adapt iteratively to new data without the need for complete retraining.

Employing the MNIST dataset, the study involved preprocessing, creating paired data for sum prediction, and implementing a train-test-validation split. A base model was developed using a convolutional neural network (CNN) on a subset of the data, which served as a baseline for further enhancements through a modified meta-learning algorithm inspired by Model-Agnostic Meta-Learning (MAML). Subsequent adaptations incorporated online meta-learning techniques to handle data arriving over time, simulating real-world scenarios where data dynamics continuously evolve.

The results demonstrated that the meta-learning models improved learning efficiency and adaptability and enhanced generalization capabilities across new data. These findings substantiate the literature on meta-learning's capacity to significantly upgrade learning processes, highlighting its practical applications across various fields. The research confirmed that meta-learning frameworks could dynamically adapt and learn from sequential data inputs, showcasing a robust improvement trajectory from the base model to the meta-model and, finally, to the online meta-model.

This study demonstrated the potential of meta-learning within artificial intelligence, particularly in enhancing machine learning models through iterative and adaptive learning. By extending the capabilities of supervised models to adapt to new and changing data swiftly without extensive retraining, meta-learning holds promise for broad applicability in more complex, real-life scenarios. Future research may explore its integration with unsupervised learning models and tackle the challenges of scalability and computational efficiency, further advancing the frontiers of machine learning technology.

Keywords: Meta-Learning, Supervised Learning, Machine Learning, Neural Networks, Deep Learning, Artificial Intelligence

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

PREFACE

This thesis is the result of my deep dive into the field of machine learning, focusing on a fascinating concept known as meta-learning or "learning to learn." It represents an academic journey and a personal challenge that I took up in October 2023 after being introduced to the topic by my supervisor, Prof. Frank Emmert-Streib. He suggested the topic "Investigation of Meta-Learning to Enhance Supervised Learning," which initially seemed daunting given my basic familiarity with machine learning and artificial intelligence, but not meta-learning. However, with determination and the invaluable support of Prof. Emmert-Streib, I embraced the challenge.

Over the past eight months, this thesis has grown from a mere idea to a comprehensive study that seeks to improve the performance of machine-learning models through meta-learning. The primary goal of my research was to find ways to optimize these models so they can learn more efficiently from labeled data, perform better, and adapt quickly to new, unseen data. Additionally, I explored how these models could continuously update themselves to remain relevant as the data they learn from changes over time.

The support I received during this time was not just academic. My parents provided constant encouragement, helping me push through challenges and focus on my goals.

This thesis is a testament to the collective effort and support from my supervisor, my family, and the numerous scholars whose works I consulted. The insights gained through this research could potentially transform the capabilities of machine learning models, making them not just learners, but efficient and adaptive learners in dynamic environments.

Tampere, 17th May 2024

Md Enamul Haque

CONTENTS

1. INTRODUCTION

1.1 Background and Context

In the dynamic landscape of machine learning, pursuing models that learn efficiently and adapt dynamically to new information is paramount [1]. Traditional machine learning approaches often struggle with tasks where data is limited or the underlying distribution changes over time. To address these challenges, meta-learning, frequently referred to as "learning to learn," has emerged as a promising approach. Meta-learning equips models with the ability to improve their learning process based on past experiences, enhancing their adaptability and performance.

Meta-learning proposes a fundamental shift in machine learning methodologies, aiming to equip algorithms with the flexibility to adapt rapidly to new tasks using prior learned knowledge [2]. The inception of meta-learning concepts can be traced back to the pioneering work by [3], who first introduced a system that could refine its learning process. This foundational perspective was advanced by the research of [4], [5], who explored meta-learning in the context of neural networks, setting the stage for subsequent developments in this field.

Significant contributions to the meta-learning landscape include the work of [6], who demonstrated the application of Long Short-Term Memory (LSTM) networks to learn optimization algorithms, moving away from the traditional hand-designed methods. This approach showcased superior performance across a range of applications, from optimizing network training protocols to image styling, highlighting the potential of learned algorithms to enhance the adaptability and efficiency of machine learning models significantly.

Further exploration by [7], who introduced Model-Agnostic Meta-Learning (MAML), a framework designed to generalize well across a wide range of tasks using minimal data points for each task. MAML's versatility has been exemplified in various applications, particularly in few-shot learning scenarios where models must adapt to new tasks rapidly with limited exposure. This framework underscores the practicality of meta-learning in real-world applications where the ability to adapt to new information quickly is crucial.

The evolution of meta-learning has also embraced online learning environments, where data arrives sequentially, and models must update continually. Harrison et al. [8] proposed Meta-Learning via Online Changepoint Analysis (MOCA), a novel approach that adapts meta-learning techniques for unsegmented time series data, allowing models to learn and adjust in real-time.

Recent studies have also highlighted meta-learning integration with deep reinforcement learning, offering pathways to enhance models' real-time decision-making capabilities. Nagabandi et al. [9] explored this integration, focusing on the continuous adaptation required in rapidly changing environments like robotics, where models must evolve as new data becomes available without resetting or retraining from scratch.

The burgeoning field of meta-learning continues to expand its horizons, addressing more complex and varied challenges. Research by [10] on Bayesian model-agnostic meta-learning (B-MAML) illustrates an advanced approach where uncertainty within tasks is managed more robustly, enhancing the model's ability to make reliable inferences in ambiguous or poorly defined scenarios.

In summary, meta-learning stands at the forefront of a paradigm shift in machine learning, promoting a transition from static, data-intensive learning methods to dynamic, adaptable, and efficient learning models. This shift is not only pivotal for the advancement of machine learning technologies but also crucial for their application in real-world settings, where adaptability and efficiency are essential. As this thesis unfolds, we will delve into how meta-learning can be specifically harnessed to refine supervised machine-learning algorithms, aiming to enhance their performance and adaptability amidst the ever-evolving landscape of data they encounter. This exploration aligns with the overarching goal of this research: to push the boundaries of what machine-learning models can achieve by leveraging the potent capabilities of meta-learning.

1.2 Problem Statement

While meta-learning has shown promise in addressing the rapid adaptation of machine learning models in few-shot learning scenarios, there remains a significant gap in its application to enhance the performance of already built machine learning models. Predominantly, studies like those involving Model-Agnostic Meta-Learning (MAML) [7], [11] have concentrated on improving adaptability with minimal exposure to new data, primarily in few-shot scenarios. However, they rarely explore systematic enhancements to the performance of already built machine learning models through continuous and extensive data interactions that characterize real-world environments. This gap is crucial in real-world applications where data not only evolves but also accumulates over time, requiring models to adapt without the need for retraining from scratch.

This thesis addresses the critical need for meta-learning frameworks that can not only quickly adapt to new tasks, as commonly explored, but also continuously enhance and refine the efficiency and performance of already built supervised machine learning models using the same datasets they were initially trained on, as well as effectively incorporating new, incoming data streams. Such advancements would significantly depart from the predominant focus on few-shot learning, providing a robust solution for real-world applications where machine learning models will remain effective as they encounter evolving data environments. This research aims to bridge this gap by developing and testing a modified meta-learning algorithm that optimizes existing models and maintains high adaptability and performance as data evolves, challenging the current boundaries of meta-learning applications.

1.3 Thesis Statement

This thesis explores how meta-learning can significantly enhance the adaptability, efficiency, and performance of already-built supervised machine-learning models. By leveraging the principles of "learning to learn," this research aims to demonstrate how meta-learning frameworks can be applied to optimize learning algorithms and improve their generalization capabilities across new tasks. The core assertion of this thesis is that meta-learning enables machine learning models to learn more efficiently from limited data and adapt to new, unforeseen tasks without extensive retraining dynamically. This adaptability is pivotal in real-world applications where data dynamics are continually changing and where traditional models often fail due to their inability to generalize beyond their initial training conditions [7], [9].

Furthermore, the thesis will explore how these meta-learning strategies can be systematically applied to ongoing and iterative learning scenarios, enhancing models' ability to update themselves in response to new information—a crucial feature for applications in autonomous driving, financial forecasting, and healthcare diagnostics. Through detailed empirical studies and theoretical analysis, this thesis aims to contribute to the machine-learning community by providing insights into the scalability and practical application of meta-learning techniques, thereby enriching the field's understanding of how advanced learning frameworks can be effectively utilized in practice.

1.4 Objectives

The primary objectives of this thesis are to:

Optimize Supervised Machine Learning Model: Investigate the application of meta-learning techniques to enhance the efficiency, performance, and adaptability of supervised learning models, facilitating faster learning and improved generalization to new data.

Adaptation to Iterative and New Data: Explore how meta-learning can enable machine learning models to dynamically update and refine their learning processes in response to new data points that come over time, thus maintaining relevance and accuracy over time without complete retraining.

These objectives aim to demonstrate the transformative potential of meta-learning in making supervised machine-learning models more robust and adaptable to the changing dynamics of real-world data.

1.5 Scope of the Study

This thesis focuses on the application of meta-learning techniques within the realm of supervised machine learning. Specifically, it examines the enhancement of learning algorithms through meta-learning frameworks such as Model-Agnostic Meta-Learning (MAML) and its variants, aiming to improve their efficiency and adaptability across new tasks. The scope includes evaluating these techniques on standard benchmark datasets like MNIST to measure improvements in learning speed and task adaptability.

A fundamental limitation of this study is its reliance on these benchmark datasets. Although it provides a controlled environment for testing and comparison, it may not fully capture the complexity or diversity of real-world data. Additionally, the study primarily addresses the adaptability and efficiency of learning models in supervised settings, with less focus on unsupervised or reinforcement learning scenarios, which could also benefit from meta-learning approaches.

Furthermore, while the thesis explores the theoretical aspects of meta-learning and conducts empirical evaluations, the computational resources required for extensive experimentation restrict the depth and breadth of the experiments performed. As such, the findings are indicative but may require further validation in larger-scale or more diverse settings to generalize the results to broader applications fully.

1.6 Significance of the Study

The findings of this thesis have the potential to impact the field of meta-learning significantly. They demonstrate the practical advantages of applying meta-learning principles to enhance the performance of supervised machine-learning models. By showcasing how these models can be optimized to adapt more swiftly and effectively to new tasks with limited data input, this research contributes to a deeper understanding of meta-learning's capacity for enhancing the flexibility and efficiency of learning algorithms.

Furthermore, the insights gained from this study are particularly relevant for advancing the implementation of meta-learning in areas such as digital twins [12], [13], where models must continuously update and adapt to new information from their physical counterparts in real time. This capability could revolutionize industries reliant on digital twins, such as manufacturing and healthcare, by enabling more accurate simulations and predictions that keep pace with changes in their real-world scenarios.

This thesis significantly enhances the meta-learning discourse by focusing on the refinement and optimization of pre-defined machine-learning models. By implementing robust meta-learning techniques beyond initial model training, this research demonstrates how established models can dynamically improve and adapt within the evolving landscapes of data and applications. This approach not only extends the functional lifespan of machine learning models but also opens new pathways for future research into sustainable, adaptable systems crucial for the progression of artificial intelligence. This focus on enhancing already-built models with meta-learning strategies addresses a critical gap in current methodologies, offering substantial improvements in model performance and efficiency in real-world applications.

1.7 Structure of the Study

This thesis is organized into seven chapters, each designed to progressively build upon the foundational concepts of meta-learning and demonstrate its applications and benefits within supervised machine learning model. Here is a brief overview of each chapter:

Introduction: This opening chapter sets the stage for the thesis by outlining the background and significance of meta-learning in the broader context of machine learning. It also defines the problem statement, thesis statement, objectives, scope, and potential impact of the research.

The goal of the Thesis: This chapter delves into the research's primary aims, focusing on enhancing supervised machine-learning models through meta-learning. It outlines specific goals, such as optimizing learning efficiency and adaptability to new data.

Literature Review: A comprehensive review of existing research in meta-learning is presented, highlighting key developments, theoretical underpinnings, and practical applications that have shaped the current landscape of meta-learning.

Methods and Data: This chapter describes the methodologies employed in the research, including the meta-learning frameworks used, the experimental setup, and the datasets involved. It also discusses the criteria for evaluating the effectiveness of the implemented meta-learning techniques.

Results: The outcomes of the empirical experiments are detailed, providing insights into how meta-learning has improved the performance and adaptability of supervised machine-learning models in various tests.

Discussion: This chapter interprets the results and discusses their implications for theory and practice. It assesses how the findings align with the thesis objectives and evaluates the broader impact of the research.

Conclusion: The final chapter summarizes the key findings of the thesis, reflects on the research contributions to the field of meta-learning, and discusses potential areas for future research. It also considers the current study's limitations and suggests avenues for further exploration.

Each chapter is crafted to ensure a logical progression, aiding the reader in understanding the complex interrelations between meta-learning and supervised learning and providing a clear path through the research conducted.

2. GOAL OF THE THESIS

2.1 Introduction

In the evolving field of machine learning, the quest for models that learn efficiently and adapt dynamically to new information is paramount. This thesis delves into the realm of meta-learning, a sophisticated approach that promises to enhance the capabilities of machine learning models beyond conventional training methods. Meta-learning, often called "learning to learn," equips models with the ability to improve their learning process based on past experiences. This chapter outlines the core objectives that guide this research, shedding light on the potential of meta-learning to transform supervised machine-learning models.

2.2 Objective 1: Optimizing Supervised Machine Learning Model

The first objective of this research is to investigate the potential of meta-learning to optimize the supervised machine-learning model. Supervised learning, the cornerstone of many machine learning applications, relies on labeled data to train models. However, these models' efficiency, performance, and adaptability can often be enhanced. This research explores how meta-learning can be applied to supervised learning models to improve these aspects. By incorporating meta-learning, we anticipate that models will learn faster and generalize better to unseen data, showcasing improved performance and efficiency.

2.3 Objective 2: Managing and Adapting to New Data Iteratively

The second objective focuses on the capability of meta-learning to manage and adapt to new data points iteratively over time. In real-world applications, data is not static; it evolves. A robust machine-learning model must adapt to these changes without complete retraining. This objective explores how meta-learning approaches can enable models to dynamically update and refine their learning in response to new information. Through iterative updates, the model can remain relevant and accurate over time, adapting to the ever-changing data landscape it encounters.

2.4 Conclusion

To conclude, this thesis aims to explore the potential of meta-learning to create machine-learning models that are efficient, effective, adaptable, and resilient to the dynamics of real-world data. These objectives pave the way for a comprehensive exploration of meta-learning's capabilities, aiming to push the boundaries of what machine-learning models can achieve.

3. LITERATURE REVIEW

3.1 Introduction to Meta-Learning

Meta-learning, often referred to as "learning to learn," offers an engaging area of study in machine learning, focusing on the development of models that adapt quickly to new tasks with limited data. This concept hinges on the idea that a model can effectively generalize to new tasks by drawing on previous experiences [14]. A groundbreaking perspective was presented by [6], who argued for a shift from manually crafting optimization algorithms to teaching models to learn these algorithms themselves. The utilization of Long Short-Term Memory (LSTM) [1] models for implementing these learned algorithms demonstrated superior efficiency over traditional, hand-designed methods in a variety of contexts, such as solving convex optimization problems, training neural networks, and styling images. This research underscores the potential benefits of adopting learned optimization strategies to enhance the adaptability and efficacy of machine learning models. Hospedales et al. [15] delivered an exhaustive examination of meta-learning, delineating its evolution and how it stands apart from traditional AI approaches that depend on static learning algorithms. They introduce a new taxonomy to classify meta-learning methods, exploring how these techniques can tackle prevalent challenges in deep learning, including scarce data, computational constraints, and the quest for enhanced generalization. Their survey delves into meta-learning's impactful applications in few-shot learning and reinforcement learning, spotlighting potential research directions and outstanding questions within the domain. Monteiro et al. [16] delved into how meta-learning can address the challenges that arise from rapid data accumulation and the need for models to adapt swiftly. They introduce a meta-learning framework designed to automate the selection of algorithms and support continuous updates without full retraining. This framework is shown to enhance the adaptability and efficiency of models in increasingly complex and dynamic data environments. Following this, Ren et al. [17] advanced the discourse on meta-learning by integrating unlabeled data with a minimal number of labelled instances for semi-supervised few-shot classification within learning episodes. They propose adaptations to Prototypical Networks, incorporating these unlabeled examples to refine class prototypes and training the model to effectively use this additional, though unlabeled, data. This methodology demonstrates improved classification performance in semi-supervised settings on adapted Omniglot and MiniImageNet benchmarks, underlining meta-learning's potential to optimize learning efficiency by harnessing labelled and unlabeled data. Ali et al. [18] introduced a novel meta-learning ensemble approach that synergizes multiple convolutional neural networks (CNNs) to elevate breast cancer classification accuracy using ultrasound images. By combining meta-learning with transfer learning and data augmentation, their approach significantly boosts feature extraction capabilities, drawing on the strengths of pre-trained networks such as Inception, ResNet50, and DenseNet121. This method not only capitalizes on the individual merits of these networks but also dynamically adapts to new and unseen data. The ensemble technique, which amalgamates outputs from various CNNs, demonstrates a marked improvement in key performance metrics such as accuracy, precision, recall, and F1 score, showcasing the model's robustness and effectiveness in medical image classification.

The landscape of meta-learning continues to be enriched by diverse and innovative contributions. For instance, Vaiciukynas et al. [19] explored a 'Two-Step Meta-Learning for Time-Series Forecasting Ensemble,' offering a fresh angle on improving forecasting accuracy. Gambetti et al. [20] delved into 'Meta-Learning Approaches for Recovery Rate Prediction,' providing valuable strategies for financial analytics. In unsupervised learning, Iwata et al. [21] presented an intriguing study on 'Meta-learning representations for clustering with infinite Gaussian mixture models,' pushing the boundaries of clustering methodologies. Qiu et al. [22] applied meta-learning to genomics, proposing a 'Meta-learning approach for genomic survival analysis' that could revolutionize personalized medicine. Retail industry forecasts get a boost from [23], who applied meta-learning for 'Retail sales forecasting,' showcasing its business applications. Lastly, Hsu et al. [24] introduced an 'Unsupervised Learning via Meta-Learning' framework, highlighting meta-learning's versatility across supervised and unsupervised domains. These contributions underscore the breadth and depth of meta-learning's impact across various scientific and practical fields.

3.2 Historical Overview of Meta-Learning

The foundational concepts of meta-learning trace back to the pioneering work of [3], who, in his groundbreaking PhD thesis, introduced a novel perspective on meta-learning. Schmidhuber emphasized the system's ability to master the art of learning itself, proposing that effective learning encompasses a variety of methods tailored to specific situations. He advocates for educational systems to focus not just on solving problems but also on enhancing the strategies for problem-solving. Highlighting the role of self-referential learning mechanisms inspired by evolutionary principles, Schmidhuber laid the groundwork for systems to evolve and refine their learning methodologies autonomously, setting a cornerstone for subsequent meta-learning research, underscoring the vital role of self-improvement and adaptability in learning systems. Following this, Bengio et al. [4]

ventured into what would later be acknowledged as meta-learning concepts, proposing a strategy to advance neural networks through the discovery of a biologically plausible synaptic learning rule. They envisaged optimizing this rule as a parametric function, applying gradient descent and genetic algorithms to bolster network performance on challenging tasks. This investigation unveiled a fresh perspective on neural network training, stressing the necessity of concurrently fine-tuning the learning rules and the network architecture inspired by biological insights, thereby illuminating a path toward more sophisticated and capable learning systems. The paper [5] delved into the specific challenges of training recurrent neural networks (RNNs), focusing on the struggle to capture long-term dependencies in input/output sequences—a critical aspect for advancing learning systems' capabilities. The paper sheds light on the limitations of gradient-based learning algorithms in maintaining long-term information, suggesting that alternative methods to traditional gradient descent might offer solutions to these challenges. This work contributes significantly to the meta-learning field by underlining the need for algorithms capable of sustaining and adapting performance over time, pivotal in the evolution of meta-learning methodologies. On a related note, Chan et al. [25] explored meta-learning within scalable data mining, demonstrating how it can effectively amalgamate various learning processes. Their research aims to craft a meta-learning framework that not only ensures high accuracy but also maintains efficiency, particularly in handling extensive databases. Their empirical study illustrates that meta-learning can match or surpass conventional learning algorithms' accuracy while providing substantial speed advantages within networked computing environments. This underscores meta-learning's potential to significantly bolster data mining applications, reinforcing its status as an effective strategy for enhancing the efficiency and accuracy of analytical processes. Thrun et al. [2] delved into the "learning to learn" concept, a cornerstone of meta-learning, discussing the pivotal role of developing systems that can incrementally improve their learning capabilities. This discussion integrates insights from various segments of machine learning and artificial intelligence, emphasizing the importance of cross-disciplinary knowledge in enhancing learning systems. Building on this foundation, Younger et al. [26] investigated the application of gradient descent methods in meta-learning, particularly with neural networks. Their work contrasts these methods with the traditionally favoured evolutionary algorithms in meta-learning contexts. They propose employing recurrent neural network (RNN) structures, including different backpropagation variants, could establish effective meta-learning frameworks. Their emphasis on Long Short-Term Memory (LSTM) networks illustrates RNNs' potential to forge learning algorithms that swiftly adapt to new tasks with limited training, marking a significant step forward in meta-learning methodologies. Further advancing the field, Hochreiter et al. [27] explored gradient descent methods to empower computers to improve or discover learning algorithms autonomously. Moving beyond evolutionary strategies' limitations, this approach leverages recurrent neural networks to facilitate meta-learning in larger, more complex systems. Their work highlights the potential of RNNs to develop sophisticated learning algorithms, particularly in predicting non-stationary time series, signifying a notable leap in applying meta-learning to more complex and parameter-dense models.

3.3 Model-Agnostic Meta-Learning (MAML)

The introduction of Model-Agnostic Meta-Learning (MAML) by [7] marked a significant milestone in meta-learning. MAML's versatility lies in its compatibility with any model that employs gradient descent for learning. It is engineered to enable rapid adaptation to new tasks with just a few gradient steps, demonstrating impressive results in few-shot learning across diverse fields such as computer vision and reinforcement learning. Expanding on MAML's foundation, Finn et al. [28] introduced a probabilistic variant to address the challenges of task ambiguity in few-shot learning. By integrating a variational parameter distribution, this extension allows for the sampling of various models to navigate ambiguous few-shot scenarios effectively. This probabilistic enhancement of MAML has been shown to bolster the flexibility of meta-learning algorithms in tackling classification and regression problems, proving particularly valuable in situations where limited data leads to vague task definitions and showcasing potential in active learning environments. Furthering the evolution of meta-learning, Yoon et al. [10] developed Bayesian model-agnostic meta-learning (B-MAML), which merges gradient-based meta-learning with non-parametric variational inference within a probabilistic framework. This innovative method enhances meta-learning's adaptability by learning complex uncertainty patterns, surpassing the standard Gaussian approximation's performance. B-MAML's approach signifies a notable advancement in refining the precision and adaptability of meta-learning models, particularly in environments where understanding and managing uncertainty is crucial. In their exploration of Bayesian Model-Agnostic Meta-Learning (B-MAML), Chen et al. [29] delved into the theoretical underpinnings to ascertain its advantages over the conventional MAML. Focusing on meta-linear regression, they illustrate that B-MAML exhibits a consistently lower risk level during meta-testing compared to MAML, thereby empirically validating B-MAML's reputed effectiveness in practice, particularly in its capability to model uncertainty with finesse. Addressing the challenge of scarce well-annotated samples in hyperspectral image (HSI) classification, [30] leverage the MAML framework. They enhance HSI classification accuracy under limited data scenarios by employing a convolutional network structure optimized through MAML, facilitating rapid task adaptability with minimal labelled data. Their method demonstrates superior performance in various HSI classification scenarios, including cross-data and cross-scene contexts, underscoring MAML's proficiency in extracting robust representations from sparse samples. Yang et al. [31] applied MAML to tackle the specific emitter identification (SEI) challenge, characterized by a paucity of labelled samples. Their adapted MAML methodology proficiently classifies electromagnetic emissions from distinct sources, achieving over 90%

accuracy in distinguishing between signals from ZigBee devices and Unmanned Aerial Vehicles (UAVs). This study highlights MAML's utility in transferring acquired knowledge to diverse tasks, significantly reducing the need for extensive retraining with fresh data.

3.4 Online Meta-Learning

Online meta-learning represents an evolution of the traditional meta-learning framework, tailored for scenarios where data is received sequentially. This adaptation emphasizes the model's ability to evolve continuously, which is crucial for applications in dynamic environments. The work of [11] is pivotal in this context. It demonstrates how meta-learning models can be fine-tuned for online settings, enhancing their flexibility and applicability in real-world situations where data and tasks evolve. In a related vein, Harrison et al. [8] introduced a cutting-edge approach known as Meta-Learning via Online Changepoint Analysis (MOCA). This method adapts meta-learning to process unsegmented time series data, a significant departure from conventional methods that rely on segmented data. By integrating a differentiable Bayesian changepoint detection mechanism, MOCA facilitates continuous meta-learning, effectively handling time series data without predefined segmentation. This methodology has proven successful in challenging environments, including nonlinear meta-regression and meta-image classification tasks, paving the way for meta-learning applications in dynamic settings where the division of functions is infeasible or unclear. These advancements collectively underscore the growing versatility of meta-learning, extending its reach to more complex and fluctuating real-world applications. Nagabandi et al. [9] delved into deep online learning through a meta-learning lens, focusing mainly on the continuous adaptation required in model-based reinforcement learning (RL). They introduce a method that marries stochastic gradient descent and an expectation-maximization algorithm with a Chinese restaurant process prior, enabling the model to adapt or evolve as it encounters shifting task distributions. This Meta-learning for Online Learning (MOLe) framework is designed to facilitate swift adjustments of deep neural network models to variable conditions, demonstrating notable proficiency in RL contexts where the environment changes, such as alterations in terrain or equipment malfunctions. On another front, Singh et al. [32] presented an innovative approach to data stream classification that merges meta-learning's efficient weight update capabilities with the dynamic nature of recurrent neural networks (RNNs), particularly in scenarios characterized by concept drift. They employed a novel meta-heuristic technique, termed optimum weight updated testing, using an enhanced opposition-based innovative updating spotted hyena optimization (ONU-SHO), which significantly boosts classification accuracy over traditional methods. This research underscores the potential of meta-learning to augment the adaptability and efficiency of streaming data classification, offering a robust solution for navigating the complexities of evolving data landscapes. Chi et al. [33] introduced an innovative approach to address the challenges of few-shot class incremental learning (FSCIL) by leveraging a meta-learning framework that employs bi-level optimization. This technique aligns training objectives closely with incremental learning tasks, utilizing bi-directional guided modulation to enhance adaptability and minimize information loss. Their method assimilates new categories with minimal data while preserving previously acquired knowledge. Demonstrating a notable edge over existing methodologies on renowned datasets such as CIFAR100, MiniImageNet, and CUB200, their approach shines in scenarios requiring the integration of new information with limited data availability, a common hurdle in incremental learning. In a similar vein, Acer et al. [34] tackled the challenge of scalability in memory usage within online meta-learning. They introduce a memory-efficient strategy that facilitates the discarding of old task instances while maintaining a dynamic state vector to encapsulate relevant information. This strategy allows for the retention of crucial insights from past tasks without the need for perfect memory, addressing a significant bottleneck in traditional meta-learning models. Their Memory Efficient Online Meta-Learning (MOML) approach, underpinned by theoretical analyses, exhibits sub-linear regret for a range of loss functions, showcasing its robustness. Empirical evaluations reveal that MOML surpasses existing methods reliant on complete recall, underscoring its utility in online learning environments where memory resources are constrained, thus offering a pragmatic solution for efficiently handling continuous streams of information.

Further contributions to online meta-learning include 'Meta-Learning Representations for Continual Learning' by [35], 'Online Learning of a Memory for Learning Rates' by [36], 'Online Structured Meta-learning' by [37], and 'Continuous Adaptation via Meta-Learning in Nonstationary and Competitive Environments' by [38] each providing unique approaches and insights that advance our understanding and implementation of learning in dynamic environments.

3.5 Conclusion

This literature review delves deeply into the multifaceted domain of meta-learning, shedding light on its core principles, evolutionary trajectory, and pivotal contributions across various application realms. It underscores meta-learning's vital role in enhancing supervised machine learning models' performance, efficiency, and adaptability, aligning seamlessly with this thesis's central aim.

The evolution of meta-learning, tracing back to the pioneering insights of [3] and extending to recent advancements by [33], has been driven by the ambition to craft systems that can self-optimize and adapt with minimal external guidance. The exploration of Model-Agnostic Meta-Learning (MAML) by [7] and its probabilistic extensions exemplifies the transformative capabilities of meta-learning, illustrating its profound impact on the adaptability of models to new and evolving scenarios.

However, despite these advancements, gaps remain in our comprehension of meta-learning's full potential, especially regarding its application in real-world, dynamic contexts and its integration in optimizing already built supervised machine learning models. This thesis aims to bridge these gaps by examining how meta-learning can further refine supervised learning models, enhancing their adaptability to adapt to new data. This research aims to contribute towards achieving this nuanced understanding, emphasizing the exploration of meta-learning's utility in real-world applications where decision-making based on prediction is paramount [39].

4. METHODS AND DATA

4.1 Introduction

This chapter explores the methodology and data that are the foundation of our study. It explores meta-learning frameworks, algorithm adaptations, data processing, model architecture and training, evaluation strategies, and computational details.

4.2 Meta-Learning Framework

In this section, we first explore two foundational meta-learning algorithms that are instrumental to our methodological framework. The first, mentioned as algorithm 1, is developed by [7], who introduced Model-Agnostic Meta-Learning (MAML). This approach is carefully constructed to enhance a model's adaptability across diverse tasks, particularly under the constraints of few-shot learning, where data is limited. MAML excels in its capacity to swiftly integrate new insights from a minimal set of examples, positioning it as a pivotal tool in our quest to refine supervised learning algorithms. This attribute aligns closely with our research's first objective, concentrating on learning faster and adapting to new tasks simultaneously.

Algorithm 1 Model-Agnostic Meta-Learning (MAML).

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters
1: Initialize θ randomly
2: **while** not done **do**
3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
4: **for all** $\mathcal{T}_i$ **do**
5: Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ with respect to K examples
6: Compute adapted parameters with gradient descent: $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$
7: **end for**
8: Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$
9: **end while**

Moving on to algorithm 2, derived from [11], this algorithm expands MAML's core principles to suit an online learning environment. It's adept at handling data that streams sequentially, facilitating the model's ongoing adaptation and eliminating the exhaustive need for retraining. This attribute aligns closely with our research's second objective, concentrating on the model's ability to dynamically adjust and update its learning process in situations where data comes over time sequentially.

Algorithm 2 Online Meta-Learning with FTML.

Input: Performance threshold of proficiency, γ
Randomly initialize w_1
Initialize the task buffer as empty, $B \leftarrow [\,]$
for $t = 1, \ldots$ **do**
  Initialize $D_t = \emptyset$
  Add $B \leftarrow B + [T_t]$
  while $|D_{T_t}| < N$ **do**
    Append batch of n new datapoints $\{(x, y)\}$ to D_t
    $w_t \leftarrow$ Meta-Update(w_t, B, t)
    $\tilde{w}_t \leftarrow$ Update-Procedure(w_t, D_t)
    if $L(D_{\text{test}}, \tilde{w}_t) < \gamma$ **then**
      Record efficiency for task T_t as $|D_t|$ datapoints
    end if
  end while
  Record final performance of $\tilde{w}_t$ on test set D_{test} for task t.
  $w_{t+1} \leftarrow w_t$
end for

Both methods represent the fundamental objective of meta-learning, which is to improve the learning process itself. By incorporating these methods into our research, we aim to enhance the performance and efficiency of machine learning models and allow them to adapt seamlessly to the always-changing data environments.

To meet our research objectives 2, we have modified the above two algorithms and developed algorithm 3 for our problem settings.

Algorithm 3 integrates meta-learning's adaptive capabilities with the structured approach of supervised learning, aiming for continuous optimization of supervised learning algorithms. Here's a breakdown of its phases:

This algorithm utilizes a convolutional neural network (CNN) as the base model to optimize performance through a structured meta-learning approach. The process begins by acquiring a combined dataset, D, divided into an initial training set, D_{initial}, and subsequent data points $D_1, D_2, \ldots, D_T$.

The learning rate for inner loops (α) and outer loops (β), epochs for base training (N), meta-updates (M), and online updates (K) are defined. A CNN model's parameters (θ) are initialized. D_{initial} is prepared through processes like pairing and preprocessing using the provided function create_pairs. The base model is trained on D_{initial} over N epochs. Performance is assessed on a validation set. The meta-learning model inherits parameters from the trained base model. Over M iterations, the model undergoes meta-

Algorithm 3 Meta-Learning for Supervised Model Optimization.

Require: Combined dataset D, divided into initial training set D_{initial} and subsequent data points $D_1, D_2, \ldots, D_T$
Require: Hyperparameters: α (learning rate for inner loop), β (learning rate for outer loop), N (number of epochs for the base model training), M (number of meta updates), K (number of online updates)
Initialize base CNN model parameters θ
Prepare D_{initial} for CNN model training (create pairs, preprocess, etc.)
Train base model on D_{initial} using standard supervised learning for N epochs
Evaluate the base model on the validation and test set and record the performance
Initialize meta-learning model with base model parameters θ
for $i = 1$ to M **do**
    Perform meta-optimization step using D_{initial} and validation set to update θ
    Evaluate and record meta-model performance on the validation set
end for
Evaluate the adapted meta-model on the test set and record performance
Initialize online meta-learning model with meta-learned parameters θ
for each subsequent data points D_t in $D_1, D_2, \ldots, D_T$ **do**
    for $j = 1$ to K **do**
        Sample task T_j from D_t
        Evaluate gradient $\nabla_\theta L(\theta, T_j)$ on T_j
        Update $\theta' = \theta - \alpha \nabla_\theta L(\theta, T_j)$
    end for
    Perform online meta-optimization step using the validation set to update θ
    Evaluate and record online meta-model performance on the validation set
end for
Test final adapted online meta-model on a separate test set

optimization using D_{initial} and validation data to update θ, evaluating performance after each update. Parameters from the meta-learning phase are used to initialize the online meta-learning model. For each data points D_t, the model performs K online updates:

- **Task Sampling:** Tasks T_j are sampled from D_t.
- **Gradient Evaluation:** Gradients $\nabla_\theta L(\theta, T_j)$ are computed.
- **Parameter Update:** Updates are made using $\theta' = \theta - \alpha \nabla_\theta L(\theta, T_j)$.

After looping through all data points, the fully adapted online meta-model is tested on a separate test set to assess its performance.

The final outputs include the adapted parameters θ, and performance metrics such as accuracy, precision, recall, and F1 score. These steps ensure the model not only learns from the initial data but also adapts continuously through online learning, improving its ability to generalize across new, unseen tasks. This comprehensive approach leverages initial deep-learning principles and advanced meta-learning techniques to create a robust model that dynamically adapts to new information, optimizing performance across diverse datasets and tasks.

4.3 Data Description and Preparation

We conducted our experiment on the MNIST dataset, a fundamental cornerstone in the machine learning landscape. It serves as the initial stepping stone for many in this field, providing a deceptively simple yet fascinating challenge. One might wonder why we should opt for such a fundamental dataset for our sophisticated meta-learning algorithm. The answer lies in MNIST's straightforwardness and transparency, making it an excellent platform to illuminate our algorithm's capabilities, focusing on its learning and adaptive capabilities without being overshadowed by data complexity. In order to ensure the reproducibility and consistency of our tests, we begin by setting a fixed random seed using the setṡeed method.

We have applied an excellent technique to pair images from the MNIST dataset to generate labels that represent the sums of the individual labels of the image pairs. This method stands in stark contrast to the typical classification task associated with MNIST, which means the model now has first to identify two digits instead of one and then predict the sum of these two digits. Let's delve into the nuances of this function and its broader implications.

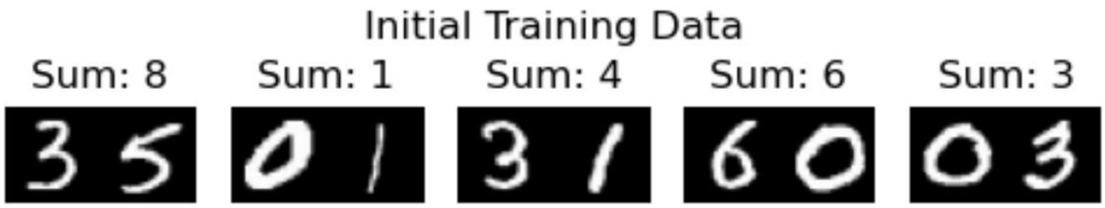

Figure 4.1. *Initial Training Data*

Figure 4.2. *Validation Data*

The process starts with the function traversing the dataset, creating pairs by aligning each image with the following one. The function replicates the grayscale image tensor across the channel dimension to adapt each image to a 3-channel format similar to an RGB image. The function then merges the two images in each pair horizontally (along the width, the second dimension in a tensor with the structure [channels, height, width]). This fusion forms a new image that encapsulates the visual data of both original images.

Figure 4.3. *Test Data*

Figure 4.4. *Data at Time Point t1*

It then calculates the sum of the labels (typically numbers between 0 and 9 in MNIST) for the images in each pair, crafting a new label. This sum acts as a composite numerical representation of the two side-by-side images. Ultimately, the function outputs a batch of these image pairs, accompanied by their cumulative labels.

Unlike the standard MNIST output, which identifies a single digit (0-9), this paired approach produces an output that's the sum of two digits, ranging from 0 (0+0) to 18 (9+9). The model's task evolves from recognizing a solitary digit to interpreting and adding two digits concurrently. The learning objective transitions from mere digit recognition to comprehending a numerical relationship (addition) between two digits. The representation of input data shifts from single-channel grayscale images to paired images, effectively expanding the width and altering the input's dimensionality. This process not only increases the difficulty level but also makes it challenging for the model to predict the target class. This approach escalates the complexity of the task, pushing the model beyond digit recognition to execute an essential arithmetic operation based on its interpretations. It signifies a move toward more sophisticated challenges that blend computer vision and arithmetic logic. It hints at the potential for multi-task learning or even a rudimentary form of visual question answering, where the model deduces the answer (the sum) from the visual cues presented.

Table 4.1 presents the shapes of the MNIST datasets at various stages of preparation. The initial training set consists of 3,819 samples, each with a shape of (3, 28, 56). The data collected at time points t1 to t10 comprises 3,818 samples, maintaining the same shape. Both the validation and test sets contain 14,000 samples each, also with a shape of (3, 28, 56). These consistent dimensions across different stages ensure uniformity in data processing and model evaluation.

Overall, the MNIST data is first divided into train, validation, and test sets. The training

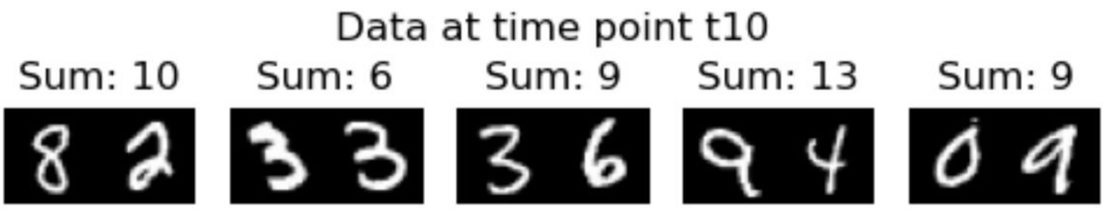

Figure 4.5. *Data at Time Point t10*

Table 4.1. *Shapes of the Datasets at Different Stages*

Stage	Shape
Initial Training Set	(3819, 3, 28, 56)
Data at Time Points $t1$ to $t10$	(3818, 3, 28, 56)
Validation Set	(14000, 3, 28, 56)
Test Set	(14000, 3, 28, 56)

data is further divided into 11 sub-sets. The first data set is mentioned as initial training data to face our first objective 2.2, and the other ten data sets are mentioned as data at time points $T_1, T_2, \ldots, T_10$, mirroring our meta-learning algorithm's iterative essence to face the second objective 2.3. Each data point signifies a distinct learning phase or time step, echoing a dynamic educational setting. The initial training data is used to train the initial base model, and the remaining 10 data points are used in the iterative process.

4.4 Model Architecture

For our experiment, we have built three models. The first model, the CNN model (mentioned as a base or initial model), is built to see how well it performs on initial training data. The second model, the meta-learning model, is built to improve the base model's performance on the same data to address the first objective 2.2. Finally, the online meta-learning model is built with 10 data points to see how the base model performs sequentially when data comes over time to address the second objective 2.3. The description of these three models can be found below:

Base Model: The base model employs a Convolutional Neural Network (CNN) designed to process images by predicting sums within a specified range (0-18). The architecture is constructed as follows:

The input layer accepts images with three channels. The first convolutional layer has 32 filters with a kernel size of 3x3 and padding of 1, followed by a ReLU activation function and a max pooling layer with a kernel size of 2. The second convolutional layer increases the filter count to 64, maintaining the same kernel size, padding, activation, and pooling configuration. Finally, a dropout layer follows with a dropout rate of 25% to prevent overfitting [40]. The output from the convolutional layers is flattened to form a single long feature vector. The first dense layer has 128 neurons, followed by a ReLU activation and a dropout rate of 50%. The output layer consists of 19 neurons (one for each class) with a softmax activation function for classification.

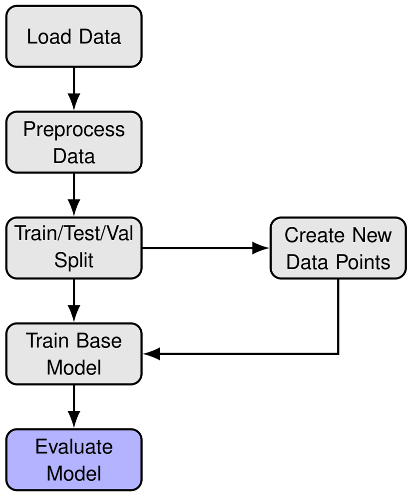

Figure 4.6. *Initial Data Processing and Base Model Training with Evaluation.*

Data is loaded and preprocessed to adjust dimensions and normalize the pixel values. It is then split into training and test sets, with the training data further divided into 11 new data points. The model is trained on the first data point, which is mentioned as initial_train_data. The model is trained for a predefined number of epochs (default is 10). Cross-entropy loss is used as the loss function, and the Adam optimizer with a learning rate of 0.001 is employed for weight updates. During training, the model's performance is continuously assessed on a validation set to monitor for improvements and potential overfitting. After training, the model is evaluated on a separate test set to measure its performance metrics, including accuracy, precision, recall, and F1-score.

The flowchart 4.6 presented in the thesis visually summarizes the data processing and training workflow. Initial data loading and preprocessing to prepare it for training. Data is divided into training, testing, and validation sets. Training data is segmented into new data points. The model is trained using the first data point. Early stopping criteria have been used to monitor overfitting in the training process with patience 5. This means if validation loss does not improve over 5 consecutive epochs, it will trigger the training. Post-training, the model is evaluated to assess its predictive performance. This systematic approach to training and evaluating the CNN ensures a robust assessment of the model's ability to generalize and perform on unseen data.

Meta-learning Model: To address objective 2.2, we explore integrating our base model, a Convolutional Neural Network (CNN), into a meta-learning framework. The base model was initially designed to handle dual-image inputs and produce three-channel data. The CNN is adapted within a meta-learning context to boost its efficacy. The model undergoes a meta-tuning process by embedding the CNN in a meta-learning framework inspired by the Model-Agnostic Meta-Learning (MAML) approach.

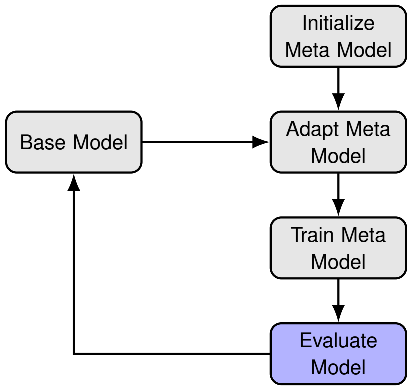

Figure 4.7. *Meta-Learning Model Training with Evaluation.*

The meta-model encapsulates the pre-trained base CNN model, using its architecture and weights as the starting point for further adaptations. Input data passes through the base model's layers, benefiting from the learned feature extraction capabilities. The meta-learning model includes an adaptation mechanism that allows for rapid learning adjustments based on new tasks, utilizing a small number of gradient updates. The meta-learning model undergoes specialized training, focusing on quick adaptation and validation. The meta-model is initialized with the weights from the trained base model to leverage previously learned patterns. The model is exposed to a new task, and it performs a series of fast adaptations (inner-loop updates) using a task-specific learning rate (lr_inner). This adaptation is conducted using Stochastic Gradient Descent (SGD) on task-specific data. Following the task-specific adaptations, the model's parameters are updated globally based on the performance across tasks (outer-loop updates). This step uses an Adam optimizer with a separate learning rate, ensuring robust generalization across a variety of functions. The meta-learning model is evaluated during and after the training process to monitor its performance and generalisation ability. After each adaptation phase, the model is evaluated on a validation set to ensure it does not overfit the specific tasks. Validation metrics such as loss and accuracy are recorded to monitor progress. Once training is complete, the model undergoes a thorough evaluation on a separate test set. This step measures the fully adapted model's performance, providing metrics like accuracy, precision, recall, and F1-score.

The meta-learning model's flowchart 4.7 illustrates the cyclical adaptation and evaluation process. It starts with the meta-model, which incorporates the base model. Then, it performs rapid on-task adaptations using the provided training data. After adaptations, the model is further trained with meta-optimization techniques. It continuously evaluates the model to refine adaptations and ensure effective learning.

This model architecture and workflow enable effective learning from limited data by quickly adapting to new tasks while maintaining the ability to generalize across new tasks. As depicted in the flowchart 4.7, the cyclical training and evaluation process ensures the model learns robustly and effectively.

Online Meta-learning Model: The OnlineMetaLearningModel is designed to explore objective 2.3, illustrating the model's ability to evolve and refine its learning processes in response to new data. This model builds upon the previously discussed CNN, which has been adapted for an online meta-learning context inspired by state-of-the-art meta-learning techniques. Leveraging the established CNN framework, the OnlineMetaLearningModel engages in successive updates, each aligned with incoming data points over time.

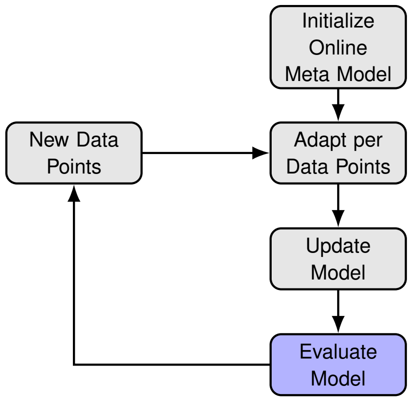

Figure 4.8. *Online Meta-Learning Model Adaptation with Evaluation.*

This process utilizes the architecture and parameters of the trained base model as the foundational framework for further learning and adaptation. It implements the forward pass through the base model, leveraging its learned features for initial data processing. The adaptive mechanism incorporates mechanisms for on-the-fly adjustments based on newly incoming data, using a specified number of updates and a learning rate tailored for the rapid integration of new insights. The training process for the online meta-learning model is characterized by its dynamic interaction with incoming data streams.

To ensure continuity in learning, the parameters of the trained base model must be considered. Then, it processes data at different time points as they arrive, performing immediate adaptations to refine the model's responses to new patterns. It utilizes Stochastic Gradient Descent (SGD) for quick updates, optimizing the model's parameters based on the latest data. After that, the model is adjusted with each new data point, employing backpropagation to minimize a loss on the latest tasks. Then, an outer-loop optimization strategy using Adam optimizer is employed to ensure the model remains robust across various data distributions. An iterative evaluation process continuously monitors the model's effectiveness and adaptability. After each update, the model is validated to assess its performance on a validation set, ensuring that new learning does not adversely affect its accuracy and stability. A comprehensive evaluation of a separate test set is conducted at

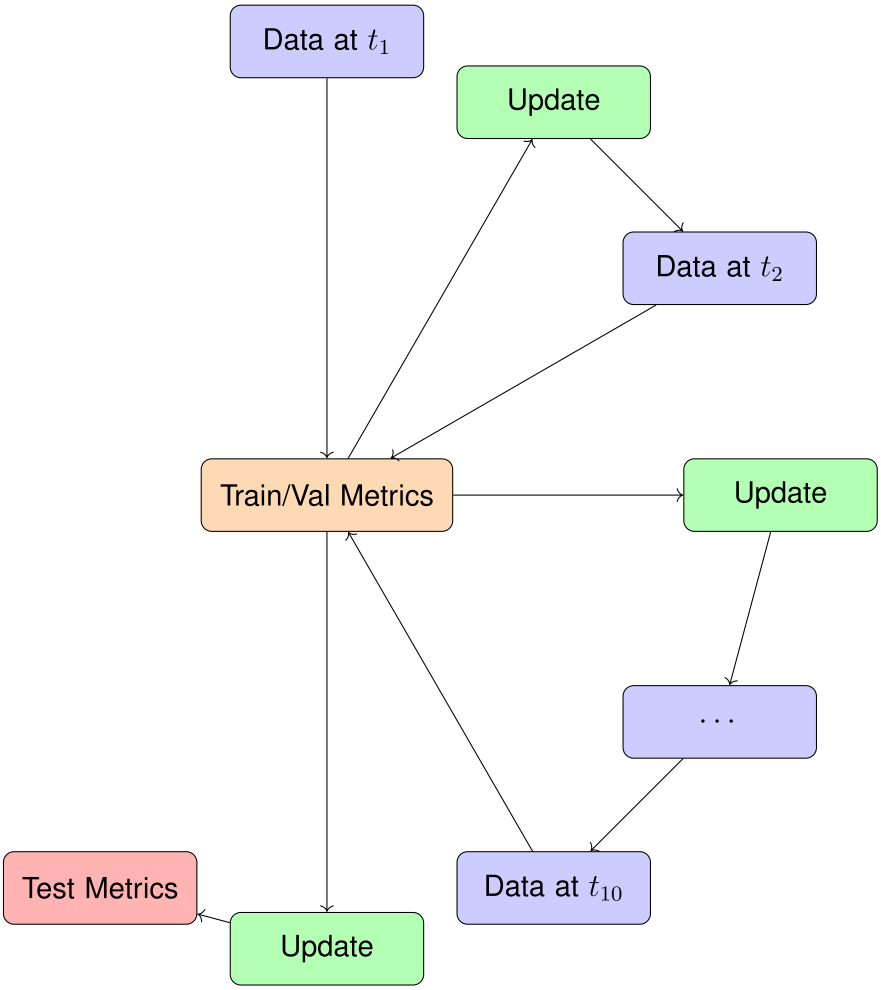

Figure 4.9. *Online Meta-Learning Model Adaptation Over Time in a Circular Layout.*

the end of the training sequence. This evaluation measures critical performance metrics such as accuracy, precision, recall, and F1-score to ensure the model's efficacy across all learned tasks.

The accompanying flowchart 4.8 elucidates the cyclic process of adaptation and evaluation inherent in the online meta-learning model. The model starts with pre-loaded base parameters and engages in rapid adaptation cycles as new data points are introduced. It regularly updates model parameters based on immediate feedback from ongoing tasks and consistently evaluates the model to ensure effective adaptation and learning from new data. The flowchart 4.9 The flow chart represents the adaptation process of an online meta-learning model over time, depicted in a circular layout. Data from sequential time points, labeled as t_1 to t_10, are used to update the model iteratively. Each dataset at a specific time point leads to an "Update" process, suggesting a continuous learning mechanism. The "Train/Val Metrics" is central to this circular flow, indicating that training and validation metrics are updated regularly as new data is processed. Additionally, "Test Metrics" are also updated, but they stem from a single point, potentially signifying a final evaluation stage after several updates. This diagram emphasizes the ongoing adaptation and evaluation of characteristics of online meta-learning frameworks that evolve based on incoming data.

This model's design facilitates continuous learning and adaptation, particularly in environments where data arrives in streams or points and requires immediate integration into the existing learning model.

4.5 Computational Details

4.5.1 Hardware Specifications

Computer: MacBook Air M1 Processor: Apple M1 chip, which includes an 8-core CPU. Memory: 8 GB RAM Storage: 256 GB ROM

4.5.2 Software and Development Environment

Operating System: macOS [Sonoma 14.4.1] Development Platform: Visual Studio Code (VS Code) equipped with the Jupyter Notebook extension.

4.5.3 Programming Language and Libraries

Programming Language: Python [3.10.13], chosen for its extensive support in data manipulation, numerical operations, and machine learning libraries. Numpy (version [1.26.3]): Used for numerical operations and handling multi-dimensional arrays. Matplotlib (version [3.8.0]): Utilized to create visualizations to analyze model performance and data characteristics. Torch (version [2.2.0]): The core library for constructing and training neural network models due to its comprehensive support for deep learning operations. Torchvision (version [0.17.0]): Provided utilities for data transformations and loading pre-defined datasets, which are crucial for preprocessing tasks. Sklearn (version [1.3.0]): Employed for splitting the data into training and testing sets, model evaluation, and performance metrics computations.

4.5.4 Implementation Details

Each neural network model was defined using PyTorch's nn.Module, and training involved leveraging automatic differentiation capabilities provided by PyTorch's torch.autograd. Model optimization was performed using the torch.optim library, with specific learning rate parameters and optimizers (e.g., Adam, SGD) selected based on the experimental design requirements. Data handling (e.g., loading, batching, preprocessing) was facilitated by a combination of torch.utils.data.DataLoader and TensorDataset, ensuring efficient memory and computational resources management during model training. Model evaluation metrics such as accuracy, precision, recall, and F1-score were computed using sklearn.metrics, providing a standardized method to assess and report model performance. Random number generation for reproducibility was controlled using random and NumPy's RNG functions, ensuring consistent and repeatable results across different runs.

This section aims to provide a detailed overview of the computational framework and tools that supported the development, training, and evaluation of the machine learning models described in the thesis. Including version numbers and specific configurations helps replicate the study and understand the computational demands of the experiments.

4.6 Summary

Overall, this chapter provided a comprehensive overview of our methodological approach, from algorithm adaptation to model training and evaluation. Through meticulous data preparation and innovative model architecture, we demonstrated the potential of meta-learning to enhance machine learning models' adaptability and efficiency. This aligned with our objectives to showcase rapid learning and dynamic adjustment in response to evolving data. The detailed computational framework ensures the reproducibility of our findings, contributing valuable insights to the field of meta-learning.

5. RESULT

5.1 Introduction

In this thesis, we wanted to determine if meta-learning can make machine-learning models better and more efficient. Think of meta-learning as a way to teach our models how to learn better. We started by looking at a base model, our starting point. This base model helps us understand what improvements meta-learning can bring. Our results section will take you through a detailed look at how our base model performed and how meta-learning added value to this process.

We first examine our base model, setting a standard for what follows. Then, we dive into more advanced models that use meta-learning. These models are designed to adapt and learn from new data better than the basic model. By comparing these models, we can see how much improvement meta-learning brings.

5.2 Base Model

The base model serves as a foundational benchmark for this study. Its performance establishes a baseline to measure the impact of meta-learning techniques on enhancing model performance.

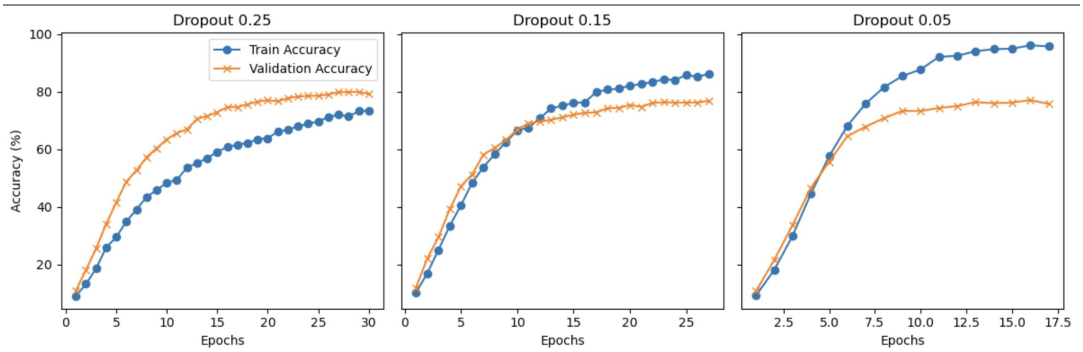

Figure 5.1. *Training and Validation Accuracy over Different Dropout Rates.*

First, we investigated the impact of different dropout rates on the base model's training dynamics and generalization performance over 30 epochs. Three CNN models were trained with dropout rates of 0.25, 0.15, and 0.05 to examine their effect on model accuracy 5.1.

The model with a dropout rate of 0.25 showed a gradual increase in both training and validation accuracy, achieving a final validation accuracy of 79.16%. In contrast, the model with a dropout rate of 0.15 demonstrated a faster improvement in accuracy, but with early stopping triggered at epoch 27, it reached a slightly lower final validation accuracy of 76.79%. The model with the lowest dropout rate of 0.05 exhibited rapid gains in training accuracy, suggesting a higher degree of fitting to the training data, but also reached early stopping at epoch 17 with a validation accuracy of 75.72%. These findings suggest that a moderate dropout rate of 0.25 provides a better balance between training performance and generalization to unseen data.

Table 5.1. *Performance of CNN Models with Different Dropout Rates on Test Data*

Dropout Rate	Test Accuracy (%)	Precision	Recall	F1 Score
0.25	78.75	0.7947	0.7875	0.7879
0.15	76.21	0.7649	0.7621	0.7620
0.05	76.75	0.7743	0.7675	0.7679

The graph 5.1 also observed that the validation accuracy exceeded the training accuracy for the model with a 0.25 dropout rate, which is an unusual scenario in machine learning. This phenomenon could be attributed to several factors. First, the dropout technique acts as a form of regularization, potentially leading the model to generalize better on the unseen validation data than on the training set. This effect is more pronounced at a higher dropout rate, such as 0.25, which helps prevent overfitting by randomly deactivating a proportion of neurons during training, thus promoting robust feature learning that generalizes well to new data [40]. Additionally, batch normalization or data augmentation techniques might improve validation performance by enhancing the model's generalisation ability.

The model with a 0.25 dropout rate also shows better results regarding test metrics 5.1. The model with a 0.25 dropout rate outperforms the other two in all test metrics like accuracy, precision, recall, and f1 score.

Choosing the model with a 0.25 dropout rate for further improvement using meta-learning is justified as it exhibits a strong balance between learning and generalization, as demonstrated by its superior performance on the validation set. This suggests that the model is not merely memorizing the training data but effectively learning transferable patterns, making it a prime candidate for further optimization through advanced techniques like meta-learning to boost its adaptability and performance on new tasks.

Table 5.2 shows the base model result. The training and evaluation of the base model across 10 epochs show a consistent improvement in training and validation loss. It indicates that the model is learning effectively from the data. Overall, the base model's training loss decreased from 2.8360 to 1.4587, and the validation loss decreased from

Table 5.2. *Base Model Training and Validation Metrics.*

Epoch	Train Loss	Val Loss	Train Acc (%)	Val Acc (%)
1/10	2.8360	2.7432	9.01	10.89
2/10	2.6339	2.4351	13.17	18.09
3/10	2.3901	2.2169	18.54	25.52
4/10	2.1544	1.9793	25.87	34.03
5/10	2.0095	1.8020	29.48	41.56
6/10	1.8532	1.6530	34.93	48.78
7/10	1.7502	1.5417	39.07	52.69
8/10	1.6366	1.4342	43.39	57.22
9/10	1.5372	1.3205	45.80	60.31
10/10	1.4587	1.2184	48.36	63.46

2.7432 to 1.2184 over 10 epochs. Correspondingly, the training accuracy improved from 9.01% to 48.36%, and the validation accuracy increased from 10.89% to 63.46%.

Table 5.3. *Base Model Test Metrics.*

Metric	Value
Test Accuracy (%)	63.19
Test Precision	0.6468
Test Recall	0.6319
Test F1 Score	0.6296

On the test set shown in table 5.3, the model achieved an accuracy of 63.19%, with a precision of 0.6468, a recall of 0.6319, and an F1 score of 0.6296.

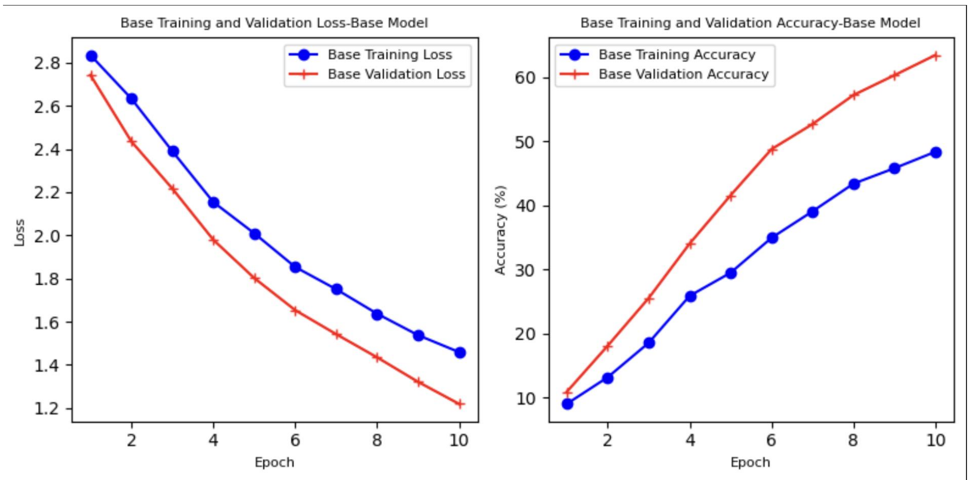

Figure 5.2. *Training and Validation Loss and Accuracy for the Base Model.*

The provided graphs, figure 5.2, for the base model, depict a positive learning trajectory over 10 epochs, with both training and validation loss decreasing significantly, which indicates effective learning. Similarly, the accuracy for both training and validation increases steadily, with validation accuracy ultimately surpassing training accuracy. This suggests the model is generalizing well rather than overfitting. It demonstrates an improvement in performance over time without compromising its ability to adapt to new, unseen data.

The consistent decrease in loss and increase in accuracy metrics over the epochs indicates that the base model is learning as expected, albeit with a relatively modest final accuracy. The difference between training and validation accuracy suggests that the model is generalizing well to unseen data, which is crucial for the model's performance in real-world scenarios.

The test set performance metrics provide a baseline for comparing the effectiveness of meta-learning-enhanced models. The test set's precision, recall, and F1 scores suggest that while the model is reasonably predictive, there is significant room for improvement, especially in balancing the precision and recall.

The results reinforce the necessity of exploring advanced techniques like meta-learning to push the boundaries of what the base model can achieve, especially regarding efficiency, adaptability, and overall performance.

Next, we will analyze the meta-model designed for Objective 2.2 to assess how meta-learning contributes to optimizing supervised learning algorithms beyond the baseline established by this base model.

5.3 Meta Model

Table 5.4 shows the meta-model's performance. The training loss remained relatively stable, fluctuating between 1.3291 and 1.3833. It suggests the model's steady adaptation during the meta-training process. The validation loss consistently decreased from 1.1335 to 0.8698, indicating the model's ability to generalize to new data with each epoch.

Table 5.4. *Meta Model Training and Validation Metrics.*

Epoch	Train Loss	Val Loss	Train Acc (%)	Val Acc (%)
1/10	1.3742	1.1335	50.85	65.81
2/10	1.3770	1.0677	50.96	67.92
3/10	1.3464	1.0312	52.45	69.22
4/10	1.3756	0.9989	51.61	70.27
5/10	1.3694	0.9767	51.58	71.04
6/10	1.3741	0.9525	52.16	71.69
7/10	1.3664	0.9281	52.08	72.63
8/10	1.3291	0.9119	52.66	73.49
9/10	1.3400	0.8911	52.61	74.02
10/10	1.3833	0.8698	51.27	75.02

The training accuracy hovered around the low 50% range, while the validation accuracy improved markedly from 65.81% to 75.02%, further reinforcing the model's generalization performance. The test evaluation metrics shown in table 5.5 indicate a substantial performance with an accuracy of 73.16%, precision at 0.7376, recall at 0.7316, and an F1 score of 0.7319.

Table 5.5. *Meta Model Test Metrics.*

Metric	**Value**
Test Loss	0.9383
Test Accuracy (%)	73.16
Test Precision	0.7376
Test Recall	0.7316
Test F1 Score	0.7319

The graph 5.3 for the meta-model shows that while the training loss remains relatively stable with minor variations, the validation loss decreases consistently across 10 epochs, suggesting the model's improving ability to generalize. The training accuracy stays relatively constant, hovering around the 50% mark, whereas the validation accuracy demonstrates a steady upward trend, ending significantly higher. This indicates that the meta-model is effectively learning from the data without overfitting, as evidenced by the gains in validation performance.

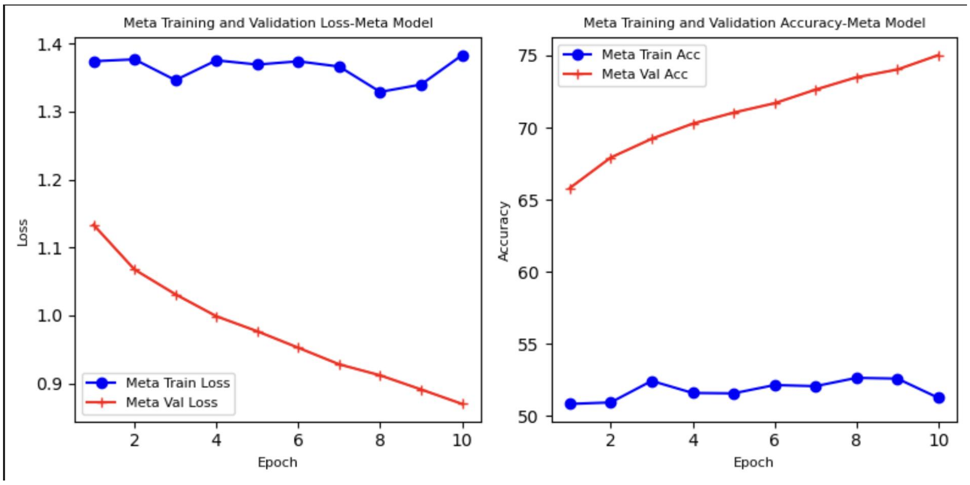

Figure 5.3. *Training and Validation Loss and Accuracy for the Meta Model.*

Despite minor fluctuations, the meta-model's training process reflects stability and resistance to overfitting, as shown by the gradual decrease in validation loss. The higher validation accuracy compared to the training accuracy by the end of the training suggests that the meta-learning algorithms have effectively enhanced the model's performance on unseen data.

The test results showcase the model's practical applicability, achieving over a 10% increase in test accuracy compared to the base model. The precision, recall, and F1 score improvements also suggest a balanced improvement across all metrics, which is vital for model reliability.

The meta-model's test performance demonstrates the efficacy of meta-learning in optimizing model adaptability and efficiency, as stated in objective 2.2 of the thesis. The enhancement in precision and recall on the test set indicates that the model is not only accurate but also consistent in its predictions across different classes. The results validate the premise that meta-learning can significantly impact the performance of supervised learning models, achieving superior results over a base model without meta-learning capabilities. The data indicates that the meta-learning model successfully met the first thesis objective, enhancing the supervised learning model's performance, efficiency, and adaptability.

Next, we will present and analyze the results of the online meta-model to understand its effectiveness in managing new points of data iteratively, addressing objective 2.3 of the thesis.

5.4 Online Meta Model

The online meta-model has been evaluated over a series of time points, reflecting its capacity to adapt iteratively.

Table 5.6. *Online Meta Model Training and Validation Metrics at Different Time Points.*

Time Point	Train Loss	Val Loss	Train Acc (%)	Val Acc (%)
t1	1.5796	1.0689	46.49	72.66
t2	1.5087	1.0113	48.95	75.00
t3	1.4847	0.9739	48.51	76.37
t4	1.4282	0.9289	50.24	77.09
t5	1.4059	0.8654	49.84	79.10
t6	1.3712	0.8512	52.41	78.76
t7	1.2909	0.7957	55.19	80.34
t8	1.3504	0.8079	53.64	80.91
t9	1.2841	0.7491	55.53	82.31
t10	1.2625	0.7363	56.50	83.00

In table 5.6, starting with an online meta-train loss of 1.5796 at time point t1, there was a steady decline to 1.2625 by t10, suggesting ongoing learning and adaptation. Online meta-validation loss reduced from 1.0689 at t1 to 0.7363 at t10, indicating an enhanced ability to generalize as the model encountered new data points. The online meta-training accuracy increased from 46.49% at the first time point to 56.50% at the tenth, while the validation accuracy saw a notable rise from 72.66% to 83.00% over the same period.

In table 5.7, the online meta model's test performance delivered impressive results: a loss of 0.7510, accuracy of 82.44%, precision at 0.8287, recall at 0.8244, and an F1 score of 0.8231.

The graphs, 5.4 depict the training and validation loss and accuracy of the online meta-

Table 5.7. *Online Meta Model Test Metrics.*

Metric	Value
Test Loss	0.7510
Test Accuracy (%)	82.44
Test Precision	0.8287
Test Recall	0.8244
Test F1 Score	0.8231

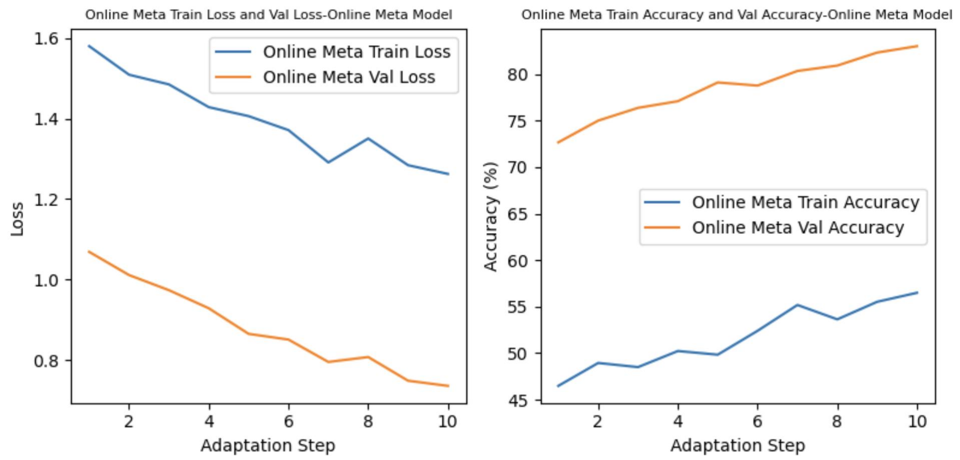

Figure 5.4. *Training and Validation Loss and Accuracy for the Online Meta Model.*

model across 10 adaptation steps. The training loss shows a fluctuating yet decreasing trend, while the validation loss decreases more smoothly, indicating effective generalization. In terms of accuracy, the training accuracy displays slight variability but trends upward, and the validation accuracy consistently increases, culminating in a notable gap where the validation accuracy exceeds the training accuracy by the final adaptation step. These trends suggest the online meta-model's increasing adeptness at handling new data over different time points.

The results reflect the online meta model's dynamic learning capability, adjusting to new information while retaining and improving upon previous knowledge. The validation accuracy surpasses the training accuracy, which typically indicates a well-generalized model that can handle unseen data effectively. The substantial increase in test metrics compared to the base model underscores the impact of iterative meta-learning techniques in enhancing model robustness and predictive performance.

These findings align with objective 2.3, demonstrating the online meta-model's effectiveness in managing and adapting to new points of data through iterative updates. The high precision and recall on the test set reflect the model's consistency and reliability across various classes of data, which is crucial for practical applications where new data streams are continuously integrated. The online meta-model successfully meets the thesis's second objective, showcasing the significant role of meta-learning in real-time data adaptability and continual learning processes. The data conclusively illustrates that the online meta-learning model adapts effectively to new data and optimizes performance over time, achieving superior outcomes compared to the base and meta models. This sets the stage for a comparative discussion to encapsulate the overall impact of meta-learning across different models.

5.5 Comparative Result

The evaluation of the base, meta, and online meta models over 10, 20, and 30 epochs demonstrates significant improvements in test metrics, including Accuracy, Precision, Recall, and F1 Score 5.5. For the base model, accuracy increased from 63.19% to 78.75%, showing steady progress across the epochs. The meta model started at a higher baseline with 73.16% accuracy at 10 epochs and improved to 84.79% by 30 epochs, reflecting more robust adaptations. Most notably, the online meta model exhibited the best performance, with initial accuracy at 82.44%, escalating to 88.33% at 30 epochs. This trend was consistent across all metrics, indicating enhanced model capability and effectiveness with additional training. These results validate the effectiveness of the training regimen and adaptations in improving model performance across different metrics.

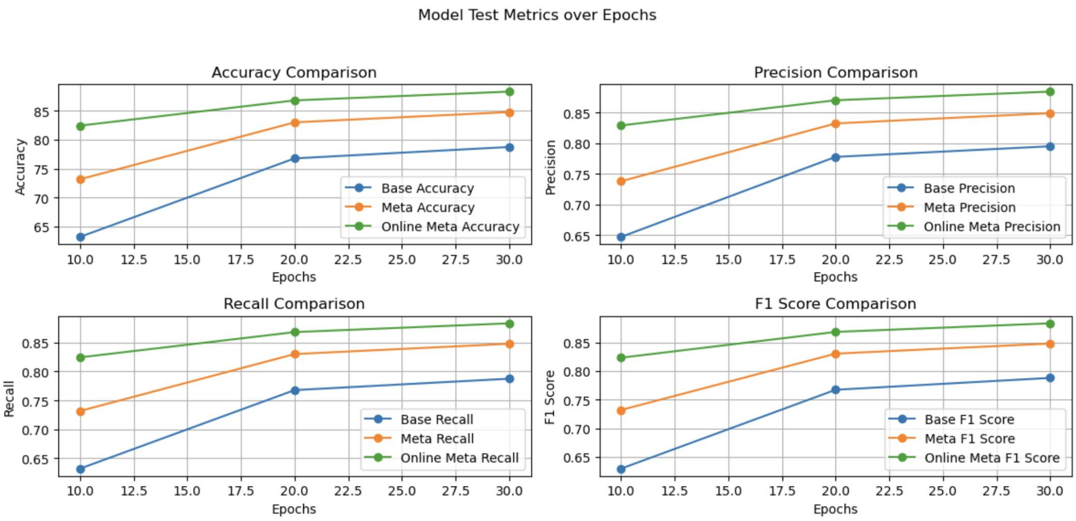

Figure 5.5. *Comparative Analysis of Test Metrics Over Different Epochs.*

The comparative graphs, 5.6 present a side-by-side evaluation of training and validation loss, alongside accuracy for the base model, the meta-model, and the online meta-model across epochs or adaptation steps.

Loss Comparison: The base model starts with the highest training and validation losses, which decrease sharply and then plateau. It suggests initial rapid learning and then stabilization. The meta model's training loss begins lower than the base model's and decreases gradually, while its validation loss reduces significantly and is the weakest among the three, suggesting efficient generalization. The online meta-model shows an

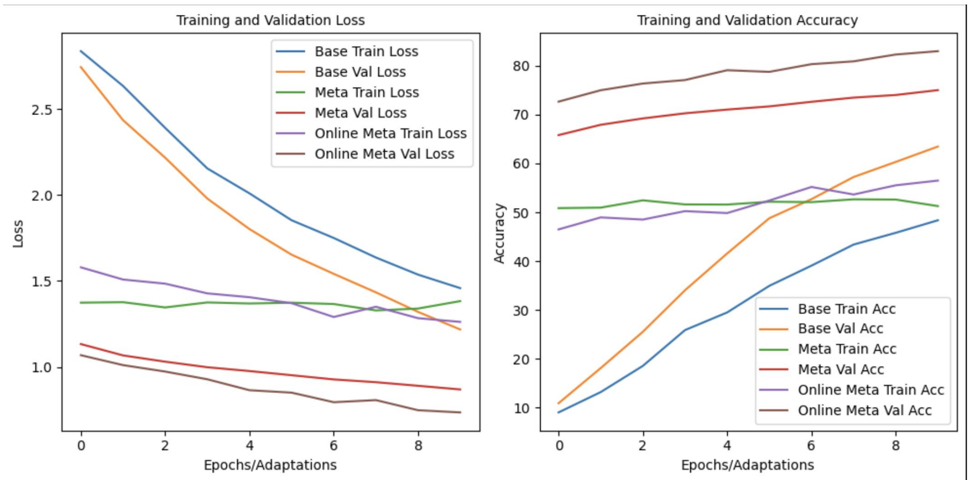

Figure 5.6. *Comparative Analysis of Training, Validation Loss, and Accuracy across Models for 10 Epochs.*

interesting pattern; its training loss starts at an intermediate level between the base and meta-models and exhibits some fluctuations. However, its validation loss consistently declines, ending lower than the base model's, which implies progressive refinement with new data.

Accuracy Comparison: For accuracy, the base model's training and validation accuracies increase slowly, with validation accuracy marginally outpacing training accuracy, demonstrating some level of generalization. The meta-model shows a substantial difference between training and validation accuracies, with the latter rising steadily to reach the highest value among the three models, indicating a solid generalization capability. The online meta-model displays an upward trend in both training and validation accuracy. Notably, the validation accuracy surpasses the base model early in the process and closes the gap with the meta-model towards later adaptations.

In summary, the base model provides a solid benchmark, while the meta model shows improved generalization. The online meta-model demonstrates the benefits of continuous adaptation, with its validation loss and accuracy approaching the meta-model's performance. This comparison does an excellent job of showing how each meta-learning approach improves supervised machine-learning models. For example, the online meta-model's continuous learning approach significantly improves over the base model.

I used 10 epochs for the base model and enhanced its performance using a meta-learning framework. Then, I investigated how it would behave if I further increased the base model's epochs. Doubling the training epochs from 10 to 20 for the base model in a meta-learning framework has demonstrably enhanced the model's performance across all evaluated metrics 5.7. The parallel reduction in validation and training loss over 20 epochs shows that the model greatly benefits from extended training. The substantial increase in training and validation accuracy demonstrates the enhanced model capability

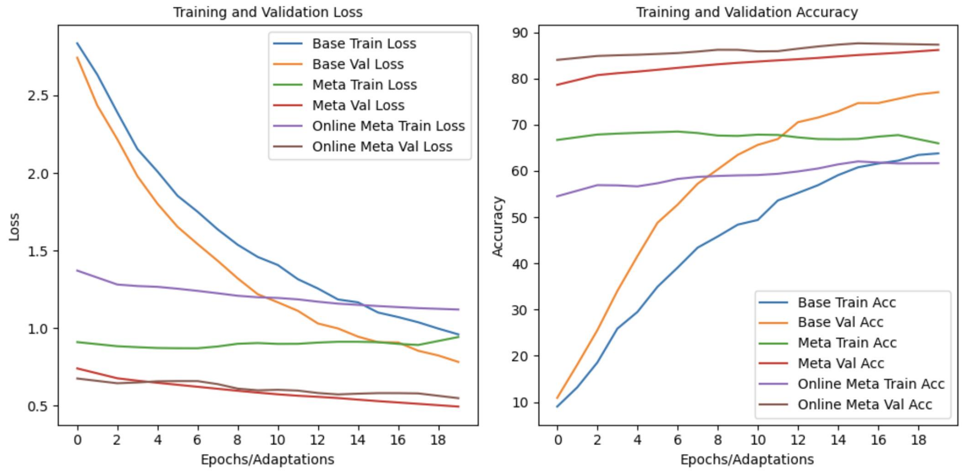

Figure 5.7. *Comparative Analysis of Training, Validation Loss, and Accuracy across Models for 20 Epochs.*

to generalize from the training data to unseen validation data. However, meta-train loss and accuracy started fluctuating after using 20 epochs in the base model.

Now, I investigated how it would behave if I further increased the epochs of the base model.

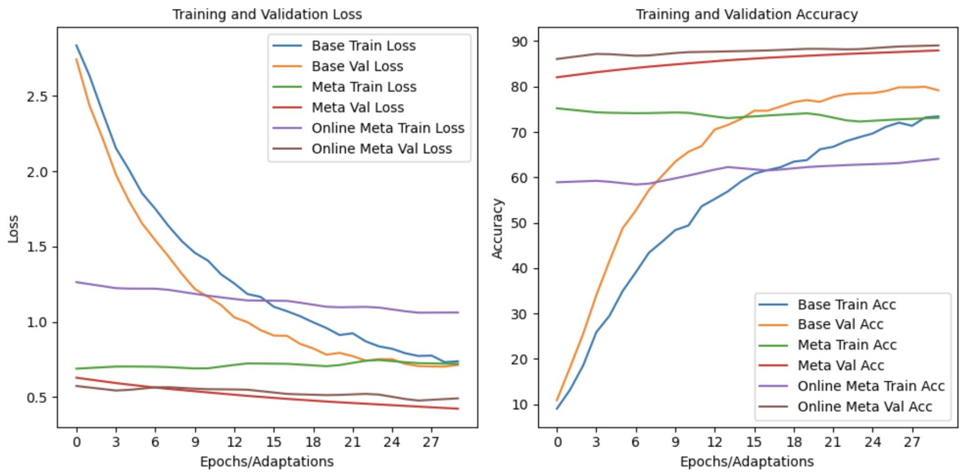

Figure 5.8. *Comparative Analysis of Training and Validation Loss and Accuracy across Models for 30 Epochs.*

The base model with 30 training epochs shows lower overall training and validation losses and higher accuracy 5.8. However, the metrics for the last ten epochs show changes, which could mean that the model is starting to learn noise and oddities from the training data instead of patterns that can be used in other situations. The meta-model that leveraged a base model trained for 30 epochs started with a lower initial validation loss and continued to improve consistently. This suggests better generalization immediately, most likely because the base model's extended training produced a more robust feature set. Similarly, the higher initial and final validation accuracy suggests that the enhanced base model provides a more effective learning foundation for the meta-model. The steady increase in accuracy also indicates more effective learning and adaptation to new data.

Overall, the difference in performance metrics (accuracy and loss) between successive blocks of epochs (1-10, 11-20, 21-30) diminishes over time. This is a common phenomenon in machine learning known as diminishing returns. Initially, when the model is underfitted, increasing the number of training epochs leads to significant improvements. However, the incremental gains from additional training decrease as the model becomes more fit to the training data. The fluctuations in training metrics, such as training accuracy and loss, could indicate that the validation metrics do not show similar improvements. As the model trains longer, it might learn the underlying patterns, noise, and specific details of the training data, which are not generalized to unseen data. As the model adapts to the training data, especially in meta-learning and online learning scenarios, the learning dynamics might fluctuate due to the complex interactions between the base model's features and the meta-model's adjustments.

In summary, based on the pattern, it's likely that further increases in epochs (say 40, 50, 60) will continue to show decreasing marginal improvements in validation and test metrics. The rate of improvement might not only decrease but potentially plateau. Continued training beyond 30 epochs might lead to overfitting unless measures such as dropout, regularization, or early stopping are employed effectively. Depending on the learning rate adjustments and other hyperparameter tuning, the fluctuations in training metrics could either stabilize if the model reaches equilibrium or worsen if the model begins to overfit more seriously.

5.6 Conclusion

At the end of our study, we looked at how our basic model, the meta-model, and the online meta-model performed. This comparison helps us see the benefits of using meta-learning in machine learning. The basic model was our starting point, showing us where to improve. We saw clear improvements when we moved to the meta-model and the online one. These improvements were in how well the models adapted to new information and how accurate they were.

The meta-model improved significantly at making predictions, showing that meta-learning can enhance a model's performance. The online meta-model was even more impressive, especially in how it adapted over time with new data.

In simple terms, our results show that meta-learning is a powerful tool for improving machine-learning models. It helped our models adapt to new situations and improve their accuracy. This chapter confirms that our study met its goals, showing that meta-learning effectively improves machine-learning models. This answers our primary questions and opens up new possibilities for future research in meta-learning.

6. DISCUSSION

In our discussion, we explore how our study contributes to knowledge about meta-learning, a concept that involves teaching machines to learn more efficiently. Our research builds upon the work of experts like [6], [14], showing how meta-learning can make machine-learning models more adaptable and efficient. We began with a base model and then enhanced it through meta-learning techniques, showing step-by-step improvements.

A significant influence on our work comes from Chelsea Finn, particularly her innovative contributions to Model-Agnostic Meta-Learning (MAML) [7] and online meta-learning [11]. We adapted her concepts to suit our unique project, employing the MNIST dataset to test our models. This adaptation not only proved the versatility of Finn's methods but also allowed us to delve into new aspects of meta-learning's potential. Through our tailored approach, we observed the strength and adaptability of meta-learning frameworks in enhancing model performance across successive data points over time, aligning with Finn's ideas about improving learning efficiency and adaptability.

Our findings support the literature's claims about meta-learning's capacity to refine learning processes, as demonstrated by the incremental improvements in our model performances. This progression aligns with theoretical discussions in the field, like those by [16]., and showcases practical applications on the MNIST dataset, highlighting meta-learning's applicability across diverse fields.

An intriguing aspect of our study was the fluctuating pattern in training loss and accuracy for the meta-learning model. While this was unexpected, it emphasized the model's rapid adaptability, a meta-learning characteristic. This was further exemplified in our online meta-learning model, which showed continual improvement, illustrating meta-learning's dynamic capabilities in evolving environments.

A critical observation in our study was the impact of different dropout rates on model performance. Dropout, as a regularization technique, plays a crucial role in preventing overfitting by randomly deactivating neurons during training. Our results indicated that a moderate dropout rate of 0.25 yielded the best balance between training accuracy and generalization to unseen data. This dropout rate not only enhanced the model's robustness but also led to a scenario where validation accuracy surpassed training accuracy, suggesting effective learning of generalizable features. These findings are consistent with

[40], who highlighted dropout's effectiveness in promoting robust feature learning and improving model generalization.

Furthermore, the dropout rate's influence on other test metrics such as precision, recall, and F1 score underscores its importance in model performance. The model with a 0.25 dropout rate outperformed others in all these metrics, suggesting that it effectively balances learning and regularization. This observation reinforces the choice of this dropout rate for further optimization using meta-learning techniques. The enhanced performance metrics indicate that the model is not merely fitting the training data but is also capable of generalizing well to new, unseen data. This robustness is crucial for deploying machine learning models in real-world applications, where data variability and novelty are constants. Our findings advocate for the strategic use of dropout in conjunction with meta-learning to build models that are not only high-performing but also resilient and adaptable.

Our research underlines meta-learning's practical implications, supporting the narrative that it holds significant potential across various sectors. The impact of our findings extends beyond theoretical discussions, providing insights into the deployment of meta-learning in different fields.

Our study paves the way for future investigations into meta-learning's scalability and broader applicability to more complex, diverse problems and real-life scenarios. Importantly, meta-learning may be relevant for digital twin research [41], [42] because the role of AI and machine learning is underdeveloped [43], [44].

In summary, our research not only underscores the theoretical advantages of meta-learning but also offers a tangible demonstration of its effectiveness in enhancing machine learning models through iterative and adaptive learning. This enriches our understanding of meta-learning's significant role in advancing the fields of artificial intelligence and machine learning, setting a solid foundation for future research in this exciting domain.

7. CONCLUSION

In this thesis, we embarked on a journey to explore the realms of meta-learning, focusing on its potential to enhance supervised machine-learning models. Through meticulous research and experimentation, we have delved into the depths of meta-learning, unravelling its capabilities and the value it adds to machine learning.

Our research had two primary goals, which served as its foundation: to investigate the role of meta-learning in optimizing supervised machine-learning models and assess how well it manages and adapts to new data points iteratively. The findings from our study significantly contributed to achieving these objectives, demonstrating that meta-learning plays a pivotal role in enhancing the adaptability, efficiency, and performance of machine learning models.

The transition from the base model to the meta-model and eventually to the online meta-model illustrated a clear trajectory of improvement, validating the hypothesis that meta-learning can be a powerful tool in refining learning algorithms. Notably, the online meta-learning model underscored the dynamic nature of meta-learning, showcasing its ability to adapt and learn from sequential data inputs continually.

While our study provides valuable insights into meta-learning's utility, it has limitations. The scope of our research was confined to specific models and the MNIST dataset, which, while offering clarity and control, also narrows the breadth of our findings' applicability. Future research could extend beyond this scope, exploring the implications of meta-learning across more diverse datasets, different model architectures, and real-world scenarios where data is vast, inherently complex, and ever-changing. Moreover, there's a vast potential for exploring how meta-learning can tackle more diverse problems, like regression tasks or applying it to unsupervised learning models.

Moreover, diving deeper into the theoretical underpinnings of meta-learning, exploring its integration with unsupervised learning models, and addressing the challenges of scalability and computational efficiency present exciting avenues for future work. Such endeavours could further unravel the potential of meta-learning, broadening its applicability and impact on the broader landscape of artificial intelligence and machine learning.

In conclusion, our study reaffirms the transformative potential of meta-learning, emphasizing its significance in advancing machine learning. As we stand on the brink of this evolving field, it is evident that the journey of meta-learning is far from over, with numerous possibilities and challenges ahead. We hope this research contributes to the existing body of knowledge and inspires further inquiry and exploration, driving forward the frontiers of meta-learning and its applications in the ever-expanding realm of artificial intelligence.

REFERENCES

[1] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer, “An introductory review of deep learning for prediction models with big data,” *Frontiers in Artificial Intelligence*, vol. 3, p. 4, 2020.

[2] S. Thrun and L. Pratt, “Learning to learn: Introduction and overview,” in *Learning to learn*, Springer, 1998, pp. 3–17.

[3] J. Schmidhuber, “Evolutionary principles in self-referential learning, or on learning how to learn: The meta-meta-... hook,” Ph.D. dissertation, Technische Universität München, 1987.

[4] Y. Bengio, S. Bengio, and J. Cloutier, *Learning a synaptic learning rule*. Citeseer, 1990.

[5] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[6] M. Andrychowicz, M. Denil, S. Gomez, *et al.*, “Learning to learn by gradient descent by gradient descent,” *Advances in neural information processing systems*, vol. 29, 2016.

[7] C. Finn, P. Abbeel, and S. Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” in *International conference on machine learning*, PMLR, 2017, pp. 1126–1135.

[8] J. Harrison, A. Sharma, C. Finn, and M. Pavone, “Continuous meta-learning without tasks,” *Advances in neural information processing systems*, vol. 33, pp. 17 571–17 581, 2020.

[9] A. Nagabandi, C. Finn, and S. Levine, “Deep online learning via meta-learning: Continual adaptation for model-based rl,” *arXiv preprint arXiv:1812.07671*, 2018.

[10] J. Yoon, T. Kim, O. Dia, S. Kim, Y. Bengio, and S. Ahn, “Bayesian model-agnostic meta-learning,” *Advances in neural information processing systems*, vol. 31, 2018.

[11] C. Finn, A. Rajeswaran, S. Kakade, and S. Levine, “Online meta-learning,” in *International conference on machine learning*, PMLR, 2019, pp. 1920–1930.

[12] F. Emmert-Streib and O. Yli-Harja, “What is a digital twin? experimental design for a data-centric machine learning perspective in health,” *International journal of molecular sciences*, vol. 23, no. 21, p. 13 149, 2022.

[13] F. Emmert-Streib, “Defining a digital twin: A data science-based unification,” *Machine Learning and Knowledge Extraction*, vol. 5, no. 3, pp. 1036–1054, 2023.

[14] R. Vilalta and Y. Drissi, “A perspective view and survey of meta-learning,” *Artificial intelligence review*, vol. 18, pp. 77–95, 2002.

[15] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, “Meta-learning in neural networks: A survey,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 9, pp. 5149–5169, 2021.

[16] J. P. Monteiro, D. Ramos, D. Carneiro, F. Duarte, J. M. Fernandes, and P. Novais, “Meta-learning and the new challenges of machine learning,” *International Journal of Intelligent Systems*, vol. 36, no. 11, pp. 6240–6272, 2021.

[17] M. Ren, E. Triantafillou, S. Ravi, *et al.*, “Meta-learning for semi-supervised few-shot classification,” *arXiv preprint arXiv:1803.00676*, 2018.

[18] M. D. Ali, A. Saleem, H. Elahi, *et al.*, “Breast cancer classification through meta-learning ensemble technique using convolution neural networks,” *Diagnostics*, vol. 13, no. 13, p. 2242, 2023.

[19] E. Vaiciukynas, P. Danenas, V. Kontrimas, and R. Butleris, “Two-step meta-learning for time-series forecasting ensemble,” *IEEE Access*, vol. 9, pp. 62 687–62 696, 2021.

[20] P. Gambetti, F. Roccazzella, and F. Vrins, “Meta-learning approaches for recovery rate prediction,” *Risks*, vol. 10, no. 6, p. 124, 2022.

[21] T. Iwata, “Meta-learning representations for clustering with infinite gaussian mixture models,” *arXiv preprint arXiv:2103.00694*, 2021.

[22] Y. L. Qiu, H. Zheng, A. Devos, H. Selby, and O. Gevaert, “A meta-learning approach for genomic survival analysis,” *Nature communications*, vol. 11, no. 1, p. 6350, 2020.

[23] S. Ma and R. Fildes, “Retail sales forecasting with meta-learning,” *European Journal of Operational Research*, vol. 288, no. 1, pp. 111–128, 2021.

[24] K. Hsu, S. Levine, and C. Finn, “Unsupervised learning via meta-learning,” *arXiv preprint arXiv:1810.02334*, 2018.

[25] P. K. Chan and S. J. Stolfo, “On the accuracy of meta-learning for scalable data mining,” *Journal of Intelligent Information Systems*, vol. 8, pp. 5–28, 1997.

[26] A. S. Younger, S. Hochreiter, and P. R. Conwell, “Meta-learning with backpropagation,” in *IJCNN’01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222)*, IEEE, vol. 3, 2001.

[27] S. Hochreiter, A. S. Younger, and P. R. Conwell, “Learning to learn using gradient descent,” in *Artificial Neural Networks—ICANN 2001: International Conference Vienna, Austria, August 21–25, 2001 Proceedings 11*, Springer, 2001, pp. 87–94.

[28] C. Finn, K. Xu, and S. Levine, “Probabilistic model-agnostic meta-learning,” *Advances in neural information processing systems*, vol. 31, 2018.

[29] L. Chen and T. Chen, “Is bayesian model-agnostic meta learning better than model-agnostic meta learning, provably?” In *International Conference on Artificial Intelligence and Statistics*, PMLR, 2022, pp. 1733–1774.

[30] K. Gao, B. Liu, X. Yu, P. Zhang, X. Tan, and Y. Sun, "Small sample classification of hyperspectral image using model-agnostic meta-learning algorithm and convolutional neural network," *International Journal of Remote Sensing*, vol. 42, no. 8, pp. 3090–3122, 2021.

[31] N. Yang, B. Zhang, G. Ding, *et al.*, "Specific emitter identification with limited samples: A model-agnostic meta-learning approach," *IEEE Communications Letters*, vol. 26, no. 2, pp. 345–349, 2021.

[32] M. N. Singh and S. Khaiyum, "Enhanced data stream classification by optimized weight updated meta-learning: Continuous learning-based on concept-drift," *International Journal of Web Information Systems*, vol. 17, no. 6, pp. 645–668, 2021.

[33] Z. Chi, L. Gu, H. Liu, Y. Wang, Y. Yu, and J. Tang, "Metafscil: A meta-learning approach for few-shot class incremental learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 14 166–14 175.

[34] D. A. E. Acar, R. Zhu, and V. Saligrama, "Memory efficient online meta learning," in *International Conference on Machine Learning*, PMLR, 2021, pp. 32–42.

[35] K. Javed and M. White, "Meta-learning representations for continual learning," *Advances in neural information processing systems*, vol. 32, 2019.

[36] F. Meier, D. Kappler, and S. Schaal, "Online learning of a memory for learning rates," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2018, pp. 2425–2432.

[37] H. Yao, Y. Zhou, M. Mahdavi, Z. J. Li, R. Socher, and C. Xiong, "Online structured meta-learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6779–6790, 2020.

[38] M. Al-Shedivat, T. Bansal, Y. Burda, I. Sutskever, I. Mordatch, and P. Abbeel, "Continuous adaptation via meta-learning in nonstationary and competitive environments," *arXiv preprint arXiv:1710.03641*, 2017.

[39] F. Emmert-Streib, S. Moutari, and M. Dehmer, "A comprehensive survey of error measures for evaluating binary decision making in data science," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 5, e1303, 2019.

[40] 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. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html.

[41] F. Emmert-Streib, "What is the role of ai for digital twins?" *AI*, vol. 4, no. 3, pp. 721–728, 2023.

[42] F. Emmert-Streib, S. Tripathi, and M. Dehmer, "Analyzing the scholarly literature of digital twin research: Trends, topics and structure," *IEEE Access*, 2023.

[43] F. Emmert-Streib, S. Moutari, and M. Dehmer, *Elements of Data Science, Machine Learning, and Artificial Intelligence Using R*. Springer Nature, 2023.

[44] F. Emmert-Streib and M. Dehmer, “Taxonomy of machine learning paradigms: A data-centric perspective,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, no. 5, e1470, 2022.