Deep Learning Methods for Mitigating Catastrophic Forgetting in Medical Imaging

by

Samaneh Javadinia
B.Sc., Shahid Beheshti University, 2014
M.Sc., Iran University of Science and Technology, 2020

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF APPLIED SCIENCE

in the Department of Electrical and Computer Engineering

University of Victoria

We acknowledge and respect the Lək̓ʷəŋən (Songhees and Xʷsepsəm/Esquimalt) Peoples on whose territory the university stands, and the Lək̓ʷəŋən and W̱SÁNEĆ Peoples whose historical relationships with the land continue to this day.

Deep Learning Methods for Mitigating Catastrophic Forgetting in Medical Imaging

by

Samaneh Javadinia
B.Sc., Shahid Beheshti University, 2014
M.Sc., Iran University of Science and Technology, 2020

Supervisory Committee

Prof. Amirali Baniasadi, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Mihai Sima, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Ashery Mbilinyi, Outside Member
(Department of Computer Science)

Supervisory Committee

Prof. Amirali Baniasadi, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Mihai Sima, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Ashery Mbilinyi, Outside Member
(Department of Computer Science)

ABSTRACT

Continual learning allows machine learning models to learn new tasks incrementally without losing previously acquired knowledge, a capability crucial in medical imaging where data evolves over time. A persistent challenge in this field is catastrophic forgetting, where models overwrite past knowledge when learning new tasks, limiting their practical use in dynamic environments. This thesis introduces a new framework called CLFCR-MC (Continual Learning Framework with Contrastive Regularization), specifically designed to tackle catastrophic forgetting in medical imaging applications. By combining momentum contrastive learning and a custom loss function that integrates classification, cosine similarity, and distillation losses, CLFCR-MC enhances the model's ability to retain previous knowledge while adapting to new tasks. Experiments using medical imaging datasets, such as BloodMNIST and PathMNIST, demonstrate that this framework significantly reduces forgetting and improves accuracy compared to existing methods. These findings highlight the potential of CLFCR-MC to address real-world challenges in continual learning and improve diagnostic capabilities in healthcare.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to **Dr. Pouya Shiri** for his invaluable support, insightful feedback, and continuous encouragement throughout my research journey. I am deeply thankful to my incredibly supportive sister, **Masumeh Javadinia**, for being my rock during the most challenging moments. Without her unwavering love and support, this work would not have been possible. I am also grateful to **my five wonderful siblings** for their constant love, encouragement, and belief in me. Each of you has contributed in your own special way to my success, and I am forever grateful for your presence in my life. I would also like to thank **my friends** for their support and companionship throughout this journey. Finally, I would like to thank **Professor Amirali Baniasadi** for his guidance and support as my supervisor.

DEDICATION

To my lovely Parents.

Chapter 1

Introduction

Medical imaging refers to the technique of creating visual representations of the interior of a body for clinical analysis and medical intervention. It includes various imaging modalities, such as X-rays, magnetic resonance imaging (MRI), and computed tomography (CT) scans, which are used to observe the structure and function of organs and tissues. The field of medical imaging plays a pivotal role in diagnosing and monitoring diseases, providing clinicians with vital information for treatment planning. In recent years, the application of machine learning, particularly deep learning models, has revolutionized medical imaging. These models have demonstrated the ability to automatically analyze medical images, performing tasks such as disease classification, image segmentation, and anomaly detection with high accuracy. In this thesis, we focus on medical imaging, specifically using machine learning to build classifiers that can interpret medical images and detect conditions based on their content. The goal is to develop accurate and reliable classifiers capable of recognizing a wide range of medical conditions from images, thereby aiding clinicians in making informed decisions. However, as new tasks and datasets emerge, a challenge known as catastrophic forgetting arises, where the model may forget previously learned knowledge while adapting to new information. To address this issue, we propose a continual learning framework that allows the model to adapt to new tasks without losing valuable knowledge from prior tasks.

1.1 Why Not Train From Ground Up?

One might argue that training a model ”from the ground up” each time new data becomes available would negate the issue of catastrophic forgetting. While this approach is technically feasible given sufficient computational resources, it is impractical and inefficient for several reasons:

- **High Computational Overhead:** Training a deep neural network from ground up on large-scale medical imaging datasets requires substantial computational resources, time, and energy. For instance, retraining a model on millions of high-resolution medical images can take several days or weeks, rendering it infeasible in environments where new data is introduced frequently.

- **Data Privacy and Storage Constraints:** Retraining requires access to the complete original dataset, which may not be permissible due to:

  - *Data privacy regulations*, such as the General Data Protection Regulation (GDPR) [35] and the Health Insurance Portability and Accountability Act (HIPAA) [29], which restrict long-term storage and reuse of sensitive medical data.
  - *Storage limitations*, as retaining large-scale datasets over extended periods incurs significant costs and is often impractical in resource-constrained settings.

    For example, medical imaging datasets such as the Cancer Imaging Archive (TCIA) [11] and UK Biobank contain hundreds of terabytes of data [39].

    A single high-resolution CT scan can range between 200MB and 2GB, depending on the resolution and number of slices [27].

    For institutions handling thousands of patients annually, this can result in petabytes of storage requirements. Additionally, cloud storage solutions such as Amazon S3 charge approximately $23 per terabyte per month [3], making long-term retention of large datasets financially burdensome.

- **Continuous Adaptation Requirements:** Clinical environments demand models that can adapt incrementally to new data without disrupting their deployment. For example, a hospital may acquire a new imaging device or encounter novel disease patterns, necessitating that the model learn new tasks while retaining prior diagnostic capabilities. Training from ground up interrupts this process and delays the deployment of updated models.

Consequently, training models ”from ground up” is neither an efficient nor a practical solution for real-world continual learning scenarios.

1.2 Why Does a Model Forget Previously Learned Tasks?

Catastrophic forgetting arises because traditional deep learning models are optimized solely for the current task, with no mechanisms to preserve prior knowledge. This phenomenon occurs due to:

- **Weight Overwriting:** Neural networks update their internal parameters to minimize the loss on the current task. When trained on new data, these updates overwrite representations associated with prior tasks, leading to the degradation of previously learned knowledge.
- **Bias Toward New Data:** Training exclusively on new data shifts the model’s focus entirely toward optimizing performance on the

current task, often at the expense of generalization to earlier tasks.

- **Lack of Explicit Memory Mechanisms:** Traditional neural networks do not explicitly retain representations of prior tasks. Unless the model has access to earlier data or a method to encode prior knowledge, it cannot effectively ”remember” previously learned tasks.

1.3 Why Prior Data or Representations Are Needed in Continual Learning

To address catastrophic forgetting, continual learning methods often incorporate mechanisms that retain access to prior data or its representations during training. These mechanisms serve to balance updates and ensure that knowledge of earlier tasks is preserved while new information is incorporated. Key reasons for using prior data or representations include:

- **Mitigating Forgetting Through Exemplar Sets:** Some continual learning methods, such as *Incremental Classifier and Representation Learning (iCaRL)* [32], utilize exemplar sets—small, representative subsets of prior data—to maintain performance on earlier tasks. For instance, storing 20 images per class from the original dataset ensures the model retains ”reminders” of previous tasks without requiring the entire dataset.

- **Knowledge Distillation:** Alternative approaches employ knowledge distillation, where a "teacher" model trained on earlier tasks transfers its knowledge to the current model by providing soft-label outputs for prior tasks. This method avoids the need to store raw data while enabling the model to retain prior knowledge.

- **Avoiding Bias Toward New Tasks:** Without access to prior data, models tend to overfit to new tasks, leading to a loss of performance on previously learned tasks. By incorporating prior data or outputs, continual learning methods balance updates and maintain generalization across tasks.

- **Practical Efficiency:** Instead of retraining the model from ground up, continual learning methods leverage compact representations of prior data, making the process more computationally efficient and scalable for dynamic, real-world environments.

Example: Imagine a continual learning model trained to diagnose both common and rare blood cell types. By retaining a small subset of images from the original dataset (e.g., 20 exemplars per class), the model can refer to these examples during training on new tasks. This ensures that it adapts to new diseases while preserving its ability to classify common blood cells.

1.4 Motivation for This Research

This thesis focuses on addressing catastrophic forgetting in the context of continual learning for medical imaging. While existing approaches, such as iCaRL, mitigate forgetting through exemplar storage and knowledge distillation, they face challenges such as representational drift (where feature embeddings shift as new tasks are introduced) and reliance on fixed exemplar sets that may not capture the complexity of evolving datasets.

Representation drift refers to the phenomenon where the feature space learned by the model shifts as new tasks are introduced. This shift in the feature space can degrade the model's performance on previously learned tasks. Specifically, as new tasks are learned, the embeddings associated with earlier tasks may be altered, making it more difficult for the model to accurately classify previously learned data.

For example, in the context of medical imaging, as new diseases or imaging modalities are introduced into the learning process, the embeddings learned for earlier diseases may change. This can result in the model being less effective at recognizing or classifying diseases it was previously trained on. For instance, if a model is initially trained to diagnose lung diseases, but new tasks related to brain imaging are introduced, the model may lose its ability to effectively classify lung diseases because the learned feature space has shifted due to the influence of the new brain imaging tasks. This shift in feature representations, or representation drift, is a significant challenge in continual learning, particularly in dynamic fields like medical imaging, where new tasks must be integrated without compromising the model's ability to perform on older tasks.

To overcome these limitations, this research proposes the *Continual Learning Framework with Contrastive Regularization via Momentum Contrast (CLFCR-MC)*. This framework combines momentum contrastive learning with a novel loss function to enhance representational stability, mitigate catastrophic forgetting, and improve adaptability to complex and evolving datasets. Experiments conducted on medical imaging datasets, such as BloodMNIST and TissueMNIST, demonstrate that CLFCR-MC achieves superior accuracy and retention compared to existing methods.

1.5 Thesis Outline

This thesis is organized as follows:

- **Chapter 2: Related Work** provides an in-depth review of state-of-the-art research on continual learning and catastrophic forgetting.
- **Chapter 3: Methodology** describes the design of the CLFCR-MC framework, detailing its architecture and loss function.

- **Chapter 4: Experimental Setup and Results** presents the datasets, evaluation metrics, and results, demonstrating the framework's effectiveness.

- **Chapter 5: Conclusion and Future Work** summarizes the research contributions and discusses potential directions for further work.

Chapter 2

Background

This chapter provides a comprehensive overview of the key concepts and methodologies in continual learning, with a focus on their relevance to addressing catastrophic forgetting in medical imaging. By exploring prominent frameworks such as Momentum Contrast (MoCo) and Incremental Classifier and Representation Learning (iCaRL), as well as the LifeLonger benchmark, this chapter lays the foundation for the proposed approach in this thesis. The discussion emphasizes the mechanisms, strengths, and limitations of these frameworks, highlighting their applications in dynamic and evolving environments like healthcare.

2.1 Continual Learning: An Overview

Continual learning is a rapidly evolving area of machine learning aimed at enabling models to learn incrementally from new data or tasks while retaining knowledge from previous ones. This approach is in stark contrast to traditional methods, which rely on training models from scratch using the entire dataset whenever new data is introduced. Continual learning is particularly critical in fields like medical imaging, where data evolves over time due to advancements in imaging technologies, the emergence of new diseases, and variations in patient demographics.

2.1.1 Challenges in Continual Learning

Despite its importance, continual learning faces several significant challenges that hinder its practical implementation:

- **Catastrophic Forgetting:** When a model is trained on new tasks, it tends to overwrite knowledge from previous tasks, leading to a loss of performance on earlier data. This problem is especially pronounced in sequential learning scenarios, where access to prior data is limited.
- **Data Constraints:** Storing and processing all past data is often impractical in domains with large datasets, such as medical imaging. Regulatory and privacy concerns, especially in healthcare, further restrict the ability to retain and reuse sensitive patient data.
- **Generalization to New Tasks:** Models must not only preserve knowledge from earlier tasks but also adapt effectively to new ones. This requires continual learning frameworks to strike a balance between *stability* (retaining old knowledge) and *plasticity* (learning new

information).

These challenges necessitate the development of innovative frameworks that mitigate forgetting, minimize resource requirements, and ensure adaptability to new tasks.

2.1.2 Relevance to Medical Imaging

In the context of medical imaging, continual learning can significantly enhance diagnostic systems. Medical datasets often grow incrementally as new imaging modalities are introduced or additional disease categories are identified. For instance, a system initially trained to classify basic blood cell types might later need to identify more specific abnormalities related to diseases such as leukemia. Continual learning enables models to incorporate such updates without sacrificing their performance on earlier tasks, ensuring their long-term utility in clinical settings.

2.2 Momentum Contrast (MoCo)

2.2.1 Overview of MoCo

Momentum Contrast (MoCo), introduced by He et al. [18], is a self-supervised learning framework designed to address challenges in learning robust and transferable feature representations. Unlike traditional supervised methods, which rely on labeled data, self-supervised learning trains models to extract meaningful features from the data itself. This approach is particularly beneficial in domains like medical imaging, where labeled datasets are often scarce and expensive to create.

MoCo's primary innovation lies in its use of a dynamic dictionary for contrastive learning. Contrastive learning is a method that teaches the model to recognize which data points are similar and which are not. For example, in medical imaging, different views of the same X-ray scan would be considered similar (positive pairs), while scans from unrelated patients would be considered dissimilar (negative pairs).

2.2.2 Mechanisms of MoCo

The architecture of MoCo consists of the following key components:

- **Query Encoder:** Processes the current input image to generate a feature representation, referred to as the *query*.
- **Momentum Encoder:** A secondary encoder that produces feature representations for past data. It evolves gradually using a moving average of the query encoder's parameters, ensuring stable and consistent representations over time.
- **Queue:** A dynamic dictionary that stores feature representations (keys) from past mini-batches. New keys are added to the queue, and the oldest keys are removed to maintain a fixed size.

The training process uses a contrastive loss function, which encourages the query to match its corresponding key (positive pair) while diverging from all other keys in the queue (negative pairs). This ensures that the model learns robust features that are both discriminative and generalizable.

2.2.3 Applications in Medical Imaging

In medical imaging, MoCo can be applied to pre-train models on large collections of unlabeled data, such as chest X-rays or CT scans. These pre-trained models can then be fine-tuned for specific tasks, such as detecting pneumonia or segmenting tumors. By leveraging self-supervised learning, MoCo reduces the dependency on labeled datasets and ensures that models are equipped with transferable features, even as new tasks are introduced.

For instance, consider a hospital deploying an AI system to analyze lung X-rays. MoCo can pre-train the system on a broad dataset of unlabeled X-rays, enabling it to learn general features of lung structures. When the hospital introduces a new diagnostic task, such as detecting COVID-19-related abnormalities, the system can be fine-tuned on a smaller labeled dataset, leveraging the pre-trained features for accurate and efficient learning.

2.2.4 Limitations and Advancements

While MoCo has demonstrated significant advantages, its reliance on a large queue of negative samples can lead to computational overhead, especially in resource-intensive domains like medical imaging. Recent advancements, such as MoCo v3, address this limitation by integrating transformer architectures and other optimizations, making the framework more scalable and efficient.

2.3 Incremental Classifier and Representation Learning (iCaRL)

2.3.1 Overview of iCaRL

Incremental Classifier and Representation Learning (iCaRL) is a framework developed for class-incremental learning, where new classes of data are introduced over time. Proposed by Rebuffi et al. [32], iCaRL aims to mitigate catastrophic forgetting by combining exemplar-based memory storage and knowledge distillation.

2.3.2 Mechanisms of iCaRL

iCaRL employs the following key techniques:

- **Exemplar Storage:** Instead of retaining all past data, iCaRL selects a small subset of representative examples, called *exemplars*, from each class. These exemplars are stored in memory and used during training to preserve knowledge of earlier classes.

- **Knowledge Distillation:** To retain prior knowledge, iCaRL aligns the outputs of the current model with those of the previous model. This ensures that the model maintains its performance on earlier tasks while adapting to new ones.

For example, in a diagnostic system trained to classify blood cell types, iCaRL would store exemplars from each blood cell category to ensure that the system continues to recognize these categories accurately, even as new cell types are introduced.

2.3.3 Applications in Medical Imaging

iCaRL is particularly suited for medical imaging scenarios where diagnostic systems encounter new disease categories over time. For instance:

- A hospital might update its diagnostic system to include new imaging modalities, such as 3D MRIs.
- iCaRL can adapt the system to classify conditions observed in the new modalities while preserving its performance on previously trained categories.

However, the reliance on exemplar storage poses challenges in privacy-sensitive domains like healthcare, where storing patient data may conflict with regulations such as GDPR or HIPAA.

2.4 LifeLonger: A Benchmark for Continual Learning in Medical Imaging

LifeLonger, introduced by Derakhshani et al. [13], is a benchmark designed to evaluate continual learning methods in medical imaging. It provides a standardized framework for assessing how well models retain knowledge while learning incrementally.

2.4.1 Key Features of LifeLonger

LifeLonger uses datasets like BloodMNIST, PathMNIST, and TissueMNIST, which represent multi-class disease classification tasks. It evaluates models in three scenarios:

- **Task Incremental Learning:** Models learn tasks sequentially, with task labels available during inference.
- **Class Incremental Learning:** Models classify all observed classes without task labels.
- **Cross-Domain Incremental Learning:** Models adapt to tasks

from different datasets or imaging modalities.

2.4.2 Relevance to Medical Imaging

LifeLonger mimics real-world challenges in medical imaging, where models must adapt to new diagnostic tasks or datasets. For instance, it tests whether a model trained on X-rays can adapt to analyzing MRIs while retaining its performance on the original task.

Chapter 3

Continual Learning in Medical Imaging

This chapter provides a detailed overview of key concepts in continual learning (CL) and contrastive learning, focusing on their relevance and applications within the domain of medical imaging. Medical imaging is a dynamic field where new diagnostic tasks, imaging modalities, and data distributions are continuously introduced, necessitating machine learning systems that can adapt incrementally without requiring complete retraining. Traditional machine learning approaches assume access to static and complete datasets during training, but these assumptions are rarely realistic in clinical settings. Continual learning addresses this limitation by enabling models to integrate new information sequentially while retaining performance on previously learned tasks. This capability is particularly valuable in medical imaging, where diagnostic systems must accommodate evolving requirements, such as emerging diseases, patient-specific variability, and advancements in imaging technology.

A fundamental challenge in continual learning, however, is the phenomenon of **catastrophic forgetting**. This occurs when models experience a loss of accuracy or degrade in performance on earlier tasks as they learn new ones. Catastrophic forgetting arises because neural networks typically overwrite previously learned parameters during training on new tasks, resulting in interference between the old and new knowledge. This issue is particularly critical in medical imaging, where even minor performance drops in previously learned diagnostic tasks can have significant consequences for patient outcomes. Addressing catastrophic forgetting is therefore essential for developing robust and scalable continual learning systems that meet the stringent accuracy and reliability demands of clinical environments. This chapter explores significant research contributions in continual learning, categorizing approaches into regularization, replay, optimization, representation, and architectural methods, and provides the foundational context for the novel methods proposed in this study.

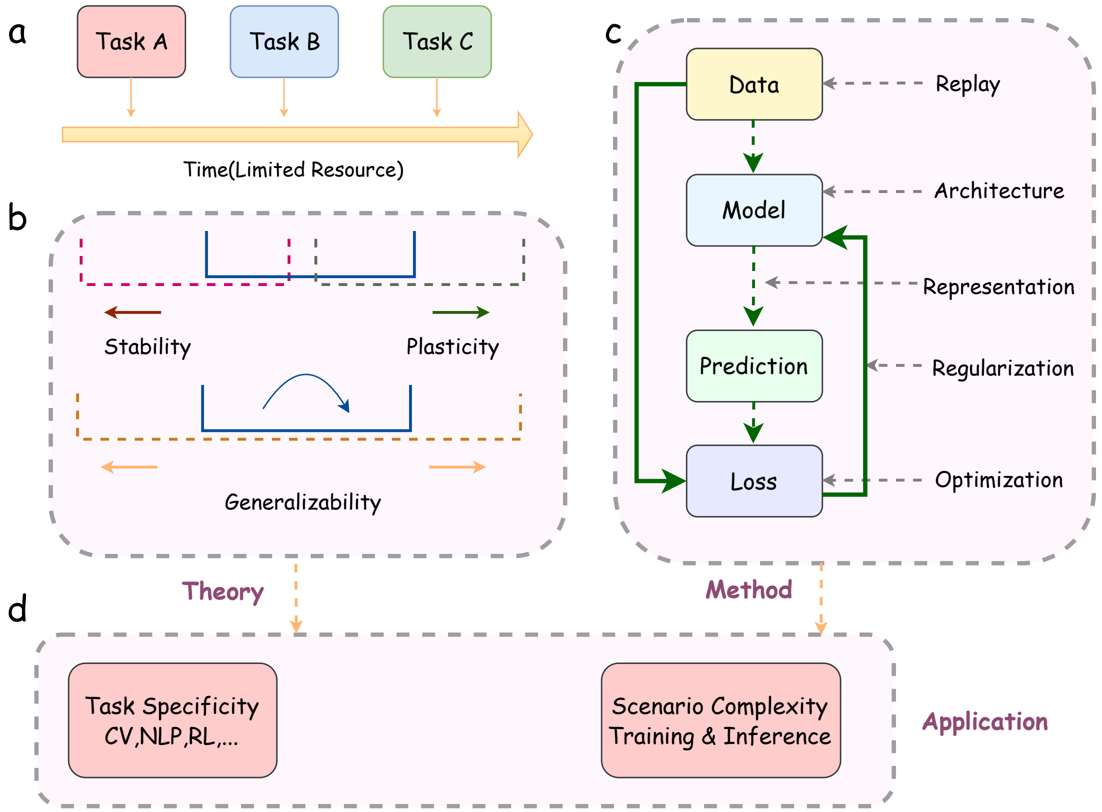


Figure 3.1: Conceptual framework of continual learning. (a) Continual learning involves incremental learning from tasks with dynamic data distributions. (b) The stability-plasticity trade-off, where excessive plasticity interferes with memory stability and vice versa. (c) Methods to address catastrophic forgetting include replay, regularization, and architectural adjustments. (d) Applications of continual learning in various machine learning domains, such as computer vision (CV) and natural language processing (NLP)[41].

This section categorizes existing approaches into regularization, replay, optimization, representation, and architectural methods, with examples of notable papers in the field.

3.1 Regularization-Based Methods

Regularization-based methods form a foundational category of approaches in continual learning that aim to mitigate catastrophic forgetting by constraining updates to the model's parameters. These methods work by introducing penalties in the loss function that discourage significant alterations to weights deemed critical for maintaining the performance of previously learned tasks. By ensuring that such parameters are preserved, these methods allow models to acquire new knowledge while retaining the ability to perform well on earlier tasks.

One of the most well-known regularization-based techniques is **Elastic Weight Consolidation (EWC)**, introduced by Kirkpatrick et al. [23]. EWC computes importance scores for the model's parameters using the Fisher Information matrix, which quantifies the sensitivity of the model's loss to changes in specific weights. During training on new tasks, EWC applies penalties that restrict updates to parameters with high importance scores, thereby preserving the knowledge encoded in those weights. This technique has proven to be particularly effective in settings with limited computational resources, as it does not require storing data from previous tasks. This advantage makes EWC well-suited for applications in medical imaging, where high-dimensional data and privacy concerns often restrict data availability.

In addition to EWC, several other regularization-based methods have been proposed, each with unique mechanisms to preserve previously acquired knowledge. Some of the most influential methods in this category include:

1. **Synaptic Intelligence (SI)** by Zenke et al. [45]: SI dynamically adjusts the importance of weights throughout the learning process. Unlike EWC, which calculates importance based on the Fisher Information matrix, SI tracks changes in parameters and their contribution to the loss reduction during training. This allows the model to adaptively identify which weights are critical to past tasks, achieving strong retention of knowledge without requiring access to previous task data. This method is particularly effective in scenarios where computational efficiency is a priority.
2. **Memory Aware Synapses (MAS)** by Aljundi et al. [2]: MAS introduces an alternative approach to identifying weight importance by estimating the relevance of parameters based on their influence on the model's outputs. Specifically, MAS measures the gradient magnitude of the output predictions with respect to each parameter, assigning higher importance to parameters with greater influence. This method is particularly advantageous in memory-constrained environments,

such as edge devices used in medical imaging applications, where storage and computational resources are limited.

3. **Learning without Forgetting (LwF)** by Li and Hoiem [26]: LwF employs a knowledge distillation strategy to preserve the model's performance on earlier tasks. During training on new tasks, the model is encouraged to produce similar outputs on old task data as it did before training. By mimicking the responses of the original model, LwF ensures that knowledge from previous tasks is retained. This method has been widely adopted in scenarios where task data is no longer accessible due to privacy constraints, making it a valuable tool in healthcare applications.

4. **Riemannian Walk** by Chaudhry et al. [6]: This method introduces a geometrically informed approach to regularization by incorporating constraints based on the curvature of the loss landscape. By penalizing updates in directions that interfere with previously learned tasks, Riemannian Walk reduces task interference and improves retention. This approach has demonstrated strong results in high-dimensional tasks, such as multi-class disease classification in medical imaging.

5. **Online EWC** by Schwarz et al. [37]: Online EWC extends the original EWC framework to accommodate streaming data, where tasks

arrive sequentially over time. This method incrementally updates importance scores in a cumulative manner, ensuring that the model remains adaptive while retaining prior knowledge. The ability to handle large-scale, continuously arriving datasets makes Online EWC particularly useful for medical imaging systems that need to integrate new patient data or imaging modalities on an ongoing basis.

Regularization-based methods are highly effective in scenarios where data from previous tasks cannot be stored or revisited, as they rely solely on the model's internal mechanisms to preserve knowledge. These methods have demonstrated significant success in medical imaging applications, such as incremental disease classification, organ segmentation across modalities, and adaptive diagnosis systems. By offering a computationally efficient and privacy-preserving approach, regularization-based techniques continue to be a cornerstone of continual learning research.

3.2 Replay-Based Methods

Replay-based methods are a prominent category in continual learning that focus on mitigating catastrophic forgetting by revisiting earlier tasks during training. This is achieved by either storing representative samples from previous tasks or generating synthetic data that can simulate earlier experiences. The core idea is to ensure that the model retains access to critical information from past tasks, thereby maintaining performance across all learned tasks. Replay-based methods have demonstrated significant potential in addressing the challenges posed by dynamic data distributions, particularly in high-stakes domains such as medical imaging.

One of the earliest and most influential examples of replay-based methods is **Incremental Classifier and Representation Learning (iCaRL)** [32]. iCaRL employs a strategy of storing a limited number of exemplars from each task, which are carefully selected to represent the overall distribution of the task data. These exemplars are then revisited during subsequent training to reinforce the model's ability to perform on earlier tasks. By balancing memory efficiency and performance, iCaRL has become a foundational approach in this category.

Replay-based methods can be broadly classified into exemplar-based approaches, which store actual data points, and generative replay approaches, which use models to recreate past data. Key works in this area include:

1. **Generative Replay** by Shin et al. [38]: Generative replay introduces a novel approach to mitigating forgetting by employing generative models to recreate past task data. Instead of storing raw data, which can raise privacy concerns and require significant memory resources, this method uses models such as Generative Adversarial Networks

(GANs) or Variational Autoencoders (VAEs) to synthesize data from previous tasks. This makes generative replay particularly suitable for privacy-sensitive domains like medical imaging, where regulations often prohibit long-term storage of patient data.

2. **Deep Generative Replay (DGR)** by Kemker and Kanan [21]: DGR builds on the concept of generative replay by integrating deep neural networks to enhance the quality and diversity of synthesized data. By generating realistic and representative samples from earlier tasks, DGR addresses the challenge of limited data availability, which is common in continual learning scenarios. This approach is especially beneficial in medical imaging applications, where datasets can be sparse, and data augmentation is critical.

3. **Exemplar-based Replay** by Rolnick et al. [34]: Exemplar-based replay involves storing a subset of representative samples from each task, which are selected based on specific criteria such as diversity or similarity to the overall data distribution. These exemplars are revisited during subsequent training to maintain the model's performance on earlier tasks. This method strikes a balance between memory efficiency and task retention, making it a practical solution for medical imaging systems that need to integrate new imaging modalities or diagnostic tasks incrementally.

4. **Memory-based Parameter Adaptation (MbPA)** by de Masson d'Autume et al. [12]: MbPA focuses on optimizing model parameters using replayed samples. By fine-tuning the model on replayed data, this approach reduces interference between tasks and ensures that knowledge from earlier tasks is retained. Although initially developed for natural language processing tasks, MbPA has significant potential for adaptation in medical imaging, where parameter tuning plays a critical role in handling multi-task scenarios.
5. **Memory Consolidation Framework** by Prabhu et al. [30]: The GDumb framework introduces an innovative approach to exemplar-based replay by creating a distilled memory buffer that stores only the most critical data points from earlier tasks. These distilled samples serve as a compact and efficient representation of previous tasks, enabling the model to achieve high retention without requiring extensive memory resources. This method is particularly relevant for edge-based medical devices, where storage and computational power are limited.

Replay-based methods are particularly valuable in medical imaging, where data from earlier tasks cannot always be stored due to privacy regulations or memory constraints. For example, in applications such as incremental disease classification, replay-based methods ensure that models retain diagnostic accuracy for previously learned conditions while adapting to newly introduced disease categories. Similarly, in multi-modal imaging systems, replay approaches help maintain consistent performance across different imaging modalities, such as CT, MRI, and ultrasound, by revisiting representative samples or synthetic data from earlier modalities.

Despite their advantages, replay-based methods also face challenges. Exemplar-based replay methods require careful selection of stored samples to avoid bias and ensure that the stored data adequately represents the original distribution. On the other hand, generative replay methods rely heavily on the quality and diversity of the synthesized data, which can vary depending on the architecture and training of the generative model. Future research in this area could focus on improving the scalability and efficiency of replay-based methods, particularly for high-dimensional medical imaging datasets, to further enhance their applicability in real-world clinical settings.

3.3 Optimization-Based Methods

Optimization-based methods are another vital category of continual learning approaches, focusing on mitigating catastrophic forgetting by controlling the updates to model parameters during training. These methods address the issue of task interference by carefully managing how gradients are propagated through the network, ensuring that the learning of new tasks does not overwrite the knowledge critical for previously learned tasks. By leveraging gradient-based constraints, optimization-based methods enable a more balanced learning process, allowing the model to retain prior knowledge while adapting to new information.

One of the foundational techniques in this category is **Gradient Episodic Memory (GEM)** [28], which adjusts the direction of gradient updates to minimize conflicts between tasks. GEM ensures that updates to model parameters align with the knowledge acquired from earlier tasks by projecting gradients into regions of the parameter space that do not interfere with previously learned information. This approach has demonstrated strong performance in high-dimensional datasets, making it particularly suitable for applications in medical imaging, where models must often handle complex, multi-modal data.

Building on the foundation of GEM, several influential studies have proposed refinements and extensions to optimization-based methods:

1. **Gradient Episodic Memory (GEM)** by Lopez-Paz and Ranzato [28]: GEM is a pioneering method in optimization-based continual learning. It projects gradient updates to ensure that the loss on previous tasks does not increase while learning new tasks. By storing a small memory of samples from earlier tasks, GEM computes task-

specific gradients and adjusts updates to minimize forgetting. This method has proven effective in scenarios involving high-dimensional datasets, such as 3D medical imaging, where preserving task-specific knowledge is crucial.

2. **A-GEM (Average GEM)** by Chaudhry et al. [7]: A-GEM refines GEM by significantly reducing memory requirements. Instead of computing task-specific gradients for each stored sample, A-GEM calculates a single representative gradient for all stored samples, streamlining the update process. This makes A-GEM particularly advantageous for resource-intensive applications in medical imaging, such as real-time diagnostics on edge devices or systems with limited computational capacity.

3. **Orthogonal Gradient Descent (OGD)** by Farajtabar et al. [14]: OGD introduces a novel approach to controlling task interference by restricting gradient updates to directions that are orthogonal to previously learned tasks. By preserving parameter space orthogonality, this method prevents overlap between tasks, ensuring better retention of earlier knowledge. OGD is highly effective in settings where tasks share overlapping feature spaces, such as the classification of related diseases using multi-modal imaging data.

4. **Meta-Experience Replay** by Riemer et al. [33]: This approach combines meta-learning principles with replay strategies to create stable updates across tasks. By incorporating meta-gradients, Meta-Experience Replay enables the model to optimize task-specific performance while simultaneously minimizing interference with previously learned tasks. This technique is particularly useful in scenarios where medical imaging models must handle long sequences of tasks, such as incremental disease classification.

5. **Learning to Learn without Forgetting** by He et al. [17]: This method introduces gradient masking as a mechanism to prevent forgetting while enabling efficient adaptation to new tasks. Gradient masking selectively blocks updates to parameters that are critical for earlier tasks, ensuring that the model retains prior knowledge. This method has been shown to be effective in complex multi-modal medical data, where features from different imaging modalities must be preserved for accurate diagnosis.

Optimization-based methods are particularly suited for medical imaging applications where task interference poses a significant challenge. For example, in scenarios where a diagnostic model is trained incrementally on data from different imaging modalities (e.g., CT and MRI) or anatomical regions (e.g., brain and chest), these methods ensure that knowledge from earlier tasks is not overwritten by subsequent updates. Furthermore, the ability of these methods to adapt to high-dimensional data makes them an ideal choice for 3D medical imaging and other computationally intensive tasks.

Despite their advantages, optimization-based methods face challenges in scaling to large datasets and long task sequences. Techniques such as GEM and A-GEM require careful memory management and gradient computation, which can become computationally expensive for extensive datasets. Future research could focus on developing more efficient gradient-based optimization strategies that balance scalability with performance retention, particularly for real-world applications in medical imaging where data diversity and task complexity are continuously increasing.

3.4 Representation-Based Methods

Representation-based methods form a crucial category in continual learning that focus on stabilizing learned representations across tasks. These approaches aim to create robust and consistent feature embeddings that remain invariant to changes in data distribution, ensuring that critical diagnostic features are preserved over time. By addressing the issue of representational drift—where feature embeddings shift as new tasks are learned—these methods play a vital role in mitigating catastrophic forgetting. This capability is particularly important in medical imaging, where maintaining stable and reliable feature representations is essential for accurate diagnosis and long-term adaptability of models.

One of the most influential representation-based methods is **Momentum Contrast (MoCo)** [18], which stabilizes representations through a dynamic memory bank and a momentum encoder. MoCo ensures that feature embeddings remain consistent over time, making it highly suitable for applications in medical imaging, where variability in imaging modalities, patient demographics, and acquisition protocols can pose significant challenges.

Several key studies have advanced the field of representation-based methods, providing innovative frameworks and techniques for improving feature stability:

1. **Momentum Contrast (MoCo)** by He et al. [18]: MoCo introduces a memory bank that stores encoded representations of previous data, ensuring consistency across tasks. The framework employs a momentum encoder that updates gradually, creating stable embeddings over time. This approach is particularly advantageous in medical imaging applications, where features must remain stable despite variability in imaging conditions or task requirements. MoCo has demonstrated strong performance in scenarios involving multi-modal imaging and

longitudinal data analysis.

2. **SimCLR** by Chen et al. [8]: SimCLR leverages a contrastive learning framework that relies on large batch sizes to create stable feature embeddings. By maximizing agreement between augmented views of the same data point while minimizing similarity with others, SimCLR generates robust representations. In medical imaging, this method is valuable for handling noisy or inconsistent datasets, as it produces embeddings that are resilient to variations in data quality.

3. **Self-Supervised Learning with Barlow Twins** by Zbontar et al. [44]: Barlow Twins focus on improving representation robustness by minimizing redundancy between feature dimensions in a self-supervised learning framework. This method offers a cost-effective alternative to traditional supervised learning, making it particularly suitable for resource-intensive domains like healthcare, where labeled data may be scarce or expensive to obtain.

4. **BYOL (Bootstrap Your Own Latent)** by Grill et al. [16]: BYOL eliminates the need for negative samples in contrastive learning by employing a bootstrapping mechanism. This method generates consistent representations by aligning outputs from two networks—one that is trained and another that evolves through a momentum-based

update. BYOL is particularly advantageous in medical imaging contexts where defining appropriate negative samples can be challenging, such as in the classification of rare diseases.

5. **SupCon (Supervised Contrastive Learning)** by Khosla et al. [22]: SupCon extends contrastive learning to supervised settings, leveraging labeled data to improve feature discrimination. By grouping similar data points closer in the embedding space and separating dissimilar ones, SupCon enhances the robustness of representations. This method is particularly effective when labeled medical datasets are available, such as annotated CT or MRI scans for specific diagnostic tasks.

Representation-based methods are particularly suited for addressing challenges in medical imaging, where consistent feature extraction across tasks is critical. For example, in applications such as disease progression monitoring, these methods ensure that the learned embeddings remain stable even as the model adapts to new imaging modalities or diagnostic categories. Similarly, in multi-modal systems, representation-based approaches facilitate the integration of features from diverse sources, such as CT, MRI, and histopathology images, into a unified diagnostic framework.

Despite their strengths, representation-based methods face challenges in scaling to large datasets and handling extreme variability in data distributions. Techniques like MoCo and BYOL rely on careful tuning of hyperparameters and require significant computational resources, which may limit their applicability in resource-constrained environments. Future research could focus on developing lightweight and scalable representation-based techniques tailored to the unique demands of medical imaging, such as handling sparse data or incorporating temporal information in dynamic imaging scenarios.

3.5 Architecture-Based Methods

Architecture-based methods represent an important category in continual learning that focuses on structural adjustments to retain knowledge and reduce task interference. These methods modify the architecture of neural networks to adapt to new tasks while preserving the knowledge acquired from previous tasks. By introducing changes to the network's structure, such as adding new neurons or pathways, these approaches aim to minimize catastrophic forgetting and maintain task-specific performance. This capability is especially critical in medical imaging, where models must handle complex, multi-task scenarios and evolving datasets.

One of the pioneering approaches in this category is **Progressive Neural Networks (PNNs)** [36], which allocates separate neurons for each task, thereby avoiding interference between tasks. PNNs provide a clear example of how architectural changes can mitigate forgetting, although they face challenges in scalability due to the increasing size of the network as new tasks are added.

Several influential studies have expanded on this concept, proposing innovative architectural methods to enhance continual learning:

1. **Progressive Neural Networks (PNNs)** by Rusu et al. [36]: PNNs create unique neural pathways for each task by freezing the parameters of earlier tasks and appending additional neurons for new tasks. This method maintains stability across tasks by preventing interference between learned and newly added parameters. While effective in preserving knowledge, PNNs require significant computational resources and memory as the network grows, which can limit their application in large-scale medical imaging systems involving numerous tasks.

2. **Dynamically Expandable Networks (DEN)** by Yoon et al. [43]: DEN addresses the scalability issue of PNNs by dynamically allocating neurons based on the complexity of the new task. Instead of adding a fixed number of neurons, DEN uses a sparsity-inducing regularizer to determine the optimal number of new parameters required for each task. This approach is particularly useful in medical imaging domains with evolving data structures, such as the addition

of new imaging modalities or disease categories, where task complexity varies significantly.

3. **PathNet** by Fernando et al. [15]: PathNet introduces a modular approach by assigning distinct neural pathways to each task. During training, the algorithm identifies and reuses relevant pathways while freezing unrelated ones, thereby minimizing parameter interference. PathNet is particularly advantageous for medical imaging applications involving overlapping features, such as the analysis of related anatomical structures or diseases across tasks.

4. **Learn to Grow (LtG)** by Li et al. [25]: LtG expands the network's capacity as needed by adding new neurons or layers for incoming tasks while selectively reusing existing parameters. By balancing flexibility and stability, LtG ensures that the network adapts efficiently to new tasks without excessive growth. This approach is particularly suitable for medical imaging systems where models must handle diverse tasks, such as disease classification, segmentation, and multi-modal data integration.

5. **Task Agnostic Continual Learning (TACO)** by Ke et al. [20]: TACO eliminates the need for task labels by modifying architectures in a task-agnostic manner. By dynamically adjusting the structure of

the network based on the data distribution, TACO enables continual learning in environments where task delineation is ambiguous. This is particularly relevant for dynamic imaging applications, such as those involving time-series medical data or unsupervised domain adaptation.

Architecture-based methods are particularly well-suited for medical imaging applications where task interference and the need for task-specific performance are significant concerns. For example, in multi-task learning scenarios such as simultaneous organ segmentation and disease classification, these methods allow the model to retain specialized parameters for each task, ensuring consistent performance across both. Similarly, in incremental learning settings where new imaging modalities or diagnostic tasks are introduced over time, architecture-based methods provide the flexibility needed to adapt to new challenges without degrading performance on earlier tasks.

Despite their advantages, architecture-based methods face challenges related to scalability and resource efficiency. Techniques such as PNNs and PathNet can lead to significant network growth as the number of tasks increases, potentially limiting their practicality in environments with constrained computational resources. Future research in this area could focus on developing more efficient architectural approaches that balance task-specific performance with scalability, particularly for resource-intensive applications in medical imaging.

3.6 Contrastive Learning in Continual Learning Frameworks

Contrastive learning, a self-supervised approach, has emerged as a powerful method for constructing robust feature representations by learning to distinguish similar and dissimilar samples in the data. This technique creates embeddings that are not only meaningful but also resilient to changes in task distribution, making it highly effective for addressing catastrophic forgetting in continual learning frameworks. By ensuring that features persist over time and remain consistent across tasks, contrastive learning plays a pivotal role in enhancing the stability of continual learning systems.

In the context of medical imaging, where data distributions can vary significantly across modalities, patient demographics, and acquisition protocols, contrastive learning enables the creation of representations that are invariant to these variations. This is particularly important for tasks such as multi-modal integration, disease classification, and longitudinal studies, where consistent embeddings are essential for accurate and reliable diagnoses.

3.6.1 Principles of Contrastive Learning

At the core of contrastive learning is the concept of learning from relationships between samples in the data. This is achieved by pairing samples into positive (similar) and negative (dissimilar) pairs and optimizing a **contrastive loss** function to bring positive pairs closer in the embedding space while pushing negative pairs apart. The contrastive loss is typically formulated as:

$$\mathcal{L}_{\text{contrastive}} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{N} \exp(\text{sim}(z_i, z_k)/\tau)}$$

where z_i and z_j are the embeddings of the positive pair, $\text{sim}(\cdot, \cdot)$ is a similarity metric such as cosine similarity, τ is a temperature parameter that controls the sharpness of the distribution, and the denominator sums over all possible pairs in the batch.

This approach allows the model to learn feature spaces that are both discriminative and invariant to noise, augmentations, or domain-specific variability. In continual learning, these stable feature spaces play a critical role in preserving knowledge across tasks, as embeddings from earlier tasks remain distinguishable even as the model adapts to new tasks.

3.6.2 Momentum Contrast (MoCo)

Momentum Contrast (MoCo) [18] extends the principles of contrastive learning by introducing a **momentum encoder** and a **queue-based memory bank**, both of which are designed to enhance the stability and scalability of representation learning. MoCo addresses some of the key limitations of traditional contrastive learning methods, such as the dependence on large batch sizes and the instability of dynamic dictionaries.

The momentum encoder in MoCo is updated using a moving average of the parameters of the query encoder, ensuring that the encoded representations in the memory bank evolve smoothly over time. This gradual update mechanism prevents abrupt changes in the feature space, which is particularly important for continual learning scenarios where tasks are introduced incrementally. Additionally, the queue-based memory bank decouples the dictionary size from the mini-batch size, enabling the model to maintain a large and diverse pool of negative samples without requiring prohibitively large batch sizes.

Key Contributions to Contrastive Learning Frameworks

MoCo has inspired several notable extensions and innovations in contrastive learning, each contributing unique advancements to the field:

1. **MoCo v2** by Chen et al. [10]: MoCo v2 refines the original MoCo

framework by incorporating stronger data augmentations, a multi-layer perceptron (MLP) head, and longer training schedules. These enhancements improve the quality and stability of embeddings, making MoCo v2 more effective for complex and high-dimensional datasets such as medical imaging data.

2. **SimSiam** by Chen and He [9]: SimSiam eliminates the need for negative samples in contrastive learning by relying on a symmetric loss function and stop-gradient operations. This framework simplifies the contrastive learning process and is particularly advantageous for applications where generating negative samples is challenging, such as in highly imbalanced medical datasets.
3. **CLIP (Contrastive Language–Image Pretraining)** by Radford et al. [31]: CLIP aligns images and text through a contrastive loss, enabling multi-modal learning. In medical imaging, CLIP provides a pathway for integrating textual clinical records with imaging data, facilitating tasks such as report generation and diagnostic explanation.
4. **Multi-Modal MoCo** by Li et al. [24]: This extension of MoCo adapts the framework for multi-modal data, allowing it to handle inputs from multiple sensors or modalities. For example, it can

integrate data from CT scans, MRIs, and ultrasound, providing a unified representation that captures complementary information from each modality.

5. **SwAV (Swapping Assignments between Views)** by Caron et al. [5]: SwAV introduces clustering techniques to contrastive learning, enabling the model to learn from data without relying on explicit pairwise comparisons. This approach is particularly useful in medical imaging scenarios with large amounts of unlabeled data, as it can discover meaningful patterns and structures without requiring manual annotations.

3.6.3 Applications in Continual Learning

Contrastive learning has the potential to play a transformative role in continual learning, particularly by enabling stable and robust feature representations that persist over time. While frameworks like MoCo and SimSiam have been successful in other contexts, their application to continual learning—especially in the domain of medical imaging—remains underexplored. This thesis pioneers the integration of contrastive learning into continual learning frameworks, leveraging its ability to create discriminative embeddings that resist representational drift.

Our approach to resisting representation drift is built upon the integration of a momentum encoder, which is updated using a momentum-based moving average method[32], specifically the Exponential Moving Average (EMA)[40]. EMA is a technique used to update parameters gradually by taking a weighted average of the previous values, with more recent values given higher weight. This method ensures that the momentum encoder evolves smoothly over time, preventing abrupt shifts in the learned feature space.

By updating the momentum encoder using this momentum-based moving average (EMA), we can ensure that its feature representations remain stable and consistent across tasks. Unlike traditional encoders that update weights directly during training, the momentum encoder's parameters evolve in a controlled manner, minimizing abrupt changes in the learned feature space. This smooth, gradual update mechanism helps reduce the risk of representation drift, a key challenge in continual learning, where the feature space can shift as new tasks are introduced.

By ensuring that the momentum encoder's parameters evolve steadily, our framework minimizes the likelihood of forgetting previously learned knowledge or degrading performance on earlier tasks. This stability in feature representations is crucial for maintaining high accuracy across multiple tasks, especially in dynamic fields like medical imaging, where consistent and reliable feature embeddings are essential for accurate diagnosis.

In the context of medical imaging, this integration provides a novel pathway to address catastrophic forgetting. For example, incremental disease classification often requires models to learn to identify new diseases without sacrificing performance on previously learned conditions. By employing contrastive learning, the embeddings for earlier tasks can remain stable, ensuring consistent performance even as the model adapts to new challenges. Additionally, in multi-modal systems, contrastive learning facilitates the integration of features from diverse modalities, such as CT, MRI, and ultrasound, while preserving task-specific knowledge.

This thesis introduces a novel framework that leverages contrastive learning in continual learning for medical imaging. By designing a tailored contrastive learning mechanism, the proposed approach addresses the limitations of existing continual learning methods, including their reliance on stored data or explicit replay mechanisms. The integration of contrastive learning ensures that feature representations remain invariant across tasks, enabling the model to efficiently retain prior knowledge while incorporating new information.

Chapter 4

Methodology

Our methodology introduces a novel and transformative approach to addressing the persistent challenge of catastrophic forgetting in continual learning frameworks. Catastrophic forgetting, a phenomenon where models lose previously learned knowledge when exposed to new tasks, significantly limits the ability of systems to handle dynamic and evolving environments. This issue is particularly critical in high-stakes domains like medical imaging, where maintaining diagnostic accuracy across changing tasks is essential.

To address catastrophic forgetting, we propose the **Continual Learning Framework with Contrastive Regularization via Momentum Contrast (CLFCR-MC)**. This framework integrates the strengths of contrastive learning with continual learning principles to ensure models retain prior knowledge while effectively adapting to new tasks. Unlike traditional methods, which often require extensive storage or replay of past data, CLFCR-MC introduces mechanisms that dynamically update feature representations over time, eliminating the reliance on large memory banks or explicit replay strategies.

The framework leverages a **Momentum Encoder**—a secondary encoder that evolves gradually using an Exponential Moving Average (EMA) update mechanism. This component provides a consistent and stable representation of past knowledge, reducing the risk of representational drift as new tasks are introduced. Additionally, the framework incorporates a novel **cosine similarity loss** that aligns representations between the main encoder and the momentum encoder, ensuring continuity in learned feature spaces.

By integrating these advancements, CLFCR-MC establishes a robust system capable of incremental learning without compromising performance on previously learned tasks. This methodology offers a significant step forward in the design of continual learning systems, particularly for applications in high-stakes domains like medical imaging, where diagnostic accuracy and the ability to handle evolving data are critical.

4.1 Harnessing Dynamic Knowledge Continuity

The core innovation of our framework lies in its ability to maintain dynamic knowledge continuity through a robust contrastive learning mechanism. At the heart of this mechanism is the integration of a **Momentum Encoder**, which is updated using an **Exponential Moving Average (EMA)**. Unlike traditional encoders that update weights directly during training, the momentum encoder evolves gradually, ensuring that its representations remain stable over time. This gradual update minimizes abrupt changes in the learned feature space, effectively preserving knowledge from previous tasks while adapting to new ones.

Our framework leverages a Momentum Encoder—a secondary encoder that evolves gradually using an Exponential Moving Average (EMA) update mechanism. This gradual update ensures that the encoder's feature representations remain stable and consistent over time. Unlike traditional methods where parameters are directly updated during training, the momentum encoder evolves without abrupt changes, minimizing the risk of representational drift as new tasks are introduced. This mechanism is key in mitigating the challenge of catastrophic forgetting, which is common in continual learning systems.

The momentum encoder plays a pivotal role in bridging the gap between current task learning and the retention of prior knowledge. By leveraging the EMA update mechanism, it generates smooth and consistent embeddings, which are less prone to the instability typically associated with continual learning. This stability is further reinforced by the cosine similarity loss, which encourages alignment between the feature representations produced by the primary encoder and the momentum encoder. Together, these components create a system that is well-suited for incremental learning tasks, particularly in medical imaging, where stable and reliable representations are crucial for diagnostic accuracy.

Building on the principles of **the Incremental Classifier and Representation Learning (iCaRL)**, our framework enhances it with a novel momentum encoder and cosine similarity loss. iCaRL relies on exemplar-based memory and distillation to retain knowledge from past tasks. However, it faces challenges such as representational drift and memory constraints. By integrating the momentum encoder with EMA and the cosine similarity loss, our framework provides a smoother and more stable feature space evolution, preserving previously learned knowledge while adapting to new tasks, even in memory-limited or privacy-sensitive applications like medical imaging.

While effective, iCaRL has inherent limitations, particularly in its handling of representational drift—the tendency for learned feature representations to shift as new tasks are introduced. This drift can lead to a degradation in the model's ability to classify data from earlier tasks accurately. Furthermore, iCaRL's reliance on stored exemplars introduces challenges related to memory constraints and privacy concerns, especially in domains like medical imaging where data is sensitive and storage is often limited by regulations.

To overcome these challenges, our proposed Continual Learning Framework with Contrastive Regularization via Momentum Contrast (CLFCR-MC) introduces two key enhancements: a **momentum encoder** and a **cosine similarity loss**. The momentum encoder, updated via an Exponential Moving Average (EMA), ensures that representations of earlier tasks remain stable and consistent over time. This mechanism effectively reduces representational drift by generating smooth and reliable embeddings that evolve gradually rather than abruptly. The cosine similarity loss further aligns these representations across tasks, encouraging the primary encoder to produce feature spaces that remain compatible with those of the momentum encoder.

These innovations not only address the limitations of iCaRL but also extend its applicability to more complex and dynamic tasks. By ensuring stable and consistent representation updates, our framework enables the model to handle a wider variety of medical imaging applications, from incremental disease classification to multi-modal integration, without compromising on accuracy or efficiency.

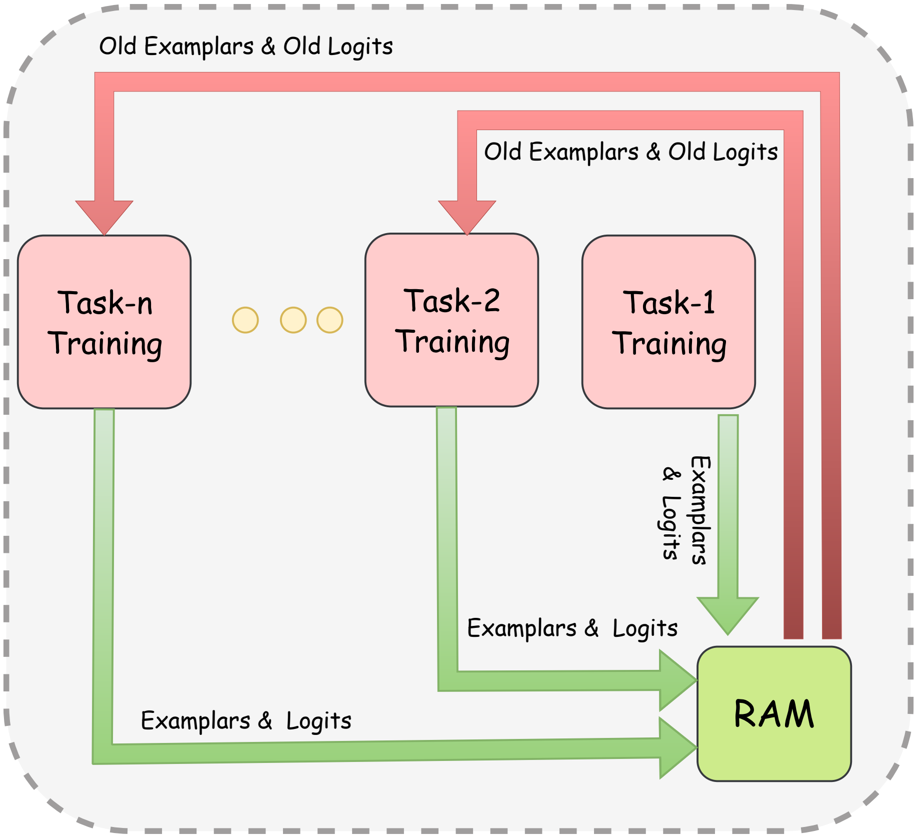


Figure 4.1: The iCaRL Architecture [32] integrates exemplar-based memory with neural adaptation to classify new tasks while preserving its previous knowledge.

In iCaRL, the model relies on exemplar-based memory to store a carefully selected subset of samples from previous tasks. This memory acts as a reference point, enabling the model to revisit past data indirectly during training on new tasks. Additionally, iCaRL incorporates a distillation loss that ensures the current model's predictions for stored exemplars remain consistent with those produced by the previous model. Together, these components help iCaRL mitigate catastrophic forgetting by balancing the retention of prior knowledge with the learning of new tasks.

While effective, iCaRL's reliance on exemplar storage and distillation loss introduces certain limitations. Exemplar-based memory can be constrained by storage capacity, which may lead to incomplete representations of past tasks, especially in scenarios with high task complexity or large datasets. Moreover, representational drift—where the learned feature space shifts as new tasks are introduced—remains a challenge, particularly for tasks requiring fine-grained diagnostic accuracy, such as those in medical imaging.

Our proposed framework, CLFCR-MC, addresses these challenges by enhancing iCaRL with two key innovations: a **momentum encoder** and a **cosine similarity loss**. The momentum encoder, updated through an Exponential Moving Average (EMA) mechanism, ensures that representations from earlier tasks evolve gradually and remain stable, effectively reducing the risk of representational drift. In parallel, the cosine similarity loss aligns the feature embeddings generated by the primary encoder with those of the momentum encoder, further enhancing consistency across tasks. These modifications significantly improve the stability of the model's learned representations, allowing it to retain knowledge from earlier tasks without sacrificing performance on new ones.

The architecture of our proposed Continual Learning Framework with Contrastive Regularization via Momentum Contrast (CLFCR-MC) builds upon the foundational principles of iCaRL while introducing significant modifications to enhance its effectiveness in handling sequential tasks. As shown in Figure 4.3, the architecture is designed to address the dual objectives of retaining knowledge from earlier tasks and efficiently adapting to new tasks. This is achieved through the integration of a **momentum encoder** and the addition of a **cosine similarity loss**, both of which work in tandem to stabilize the learning process and mitigate catastrophic forgetting.

In our framework, Task-1 represents the initial learning phase, where the model is trained on a specific dataset using both classification and distillation losses. During this phase, the momentum encoder is initialized and begins capturing representations using an Exponential Moving Average (EMA) of the primary encoder's weights. This initialization ensures that the momentum encoder provides a consistent and stable baseline for representation alignment.

As new tasks are introduced (Task-n learning), the model must simultaneously learn from current data while retaining knowledge of prior tasks. The momentum encoder plays a critical role in this phase by maintaining a smoothed version of the feature representations learned over time. By updating gradually through the EMA mechanism, the momentum encoder avoids the abrupt representational shifts that often lead to catastrophic forgetting. Additionally, the cosine similarity loss aligns the feature spaces produced by the primary encoder with those of the momentum encoder, ensuring that the learned representations remain compatible across tasks.

A key advantage of this architecture is its ability to handle incremental learning scenarios without requiring extensive storage of past data. Unlike traditional replay-based methods, which rely on storing and revisiting large amounts of previous task data, the CLFCR-MC framework uses the momentum encoder and cosine similarity loss to retain essential knowledge implicitly. This design is particularly advantageous in medical imaging, where data storage is often constrained by privacy regulations and ethical considerations.

The overall workflow of our framework ensures a seamless transition between tasks while maintaining high levels of accuracy on previously learned tasks. By dynamically updating representations and aligning feature spaces, the CLFCR-MC architecture provides a robust and scalable solution for continual learning in dynamic and evolving environments.

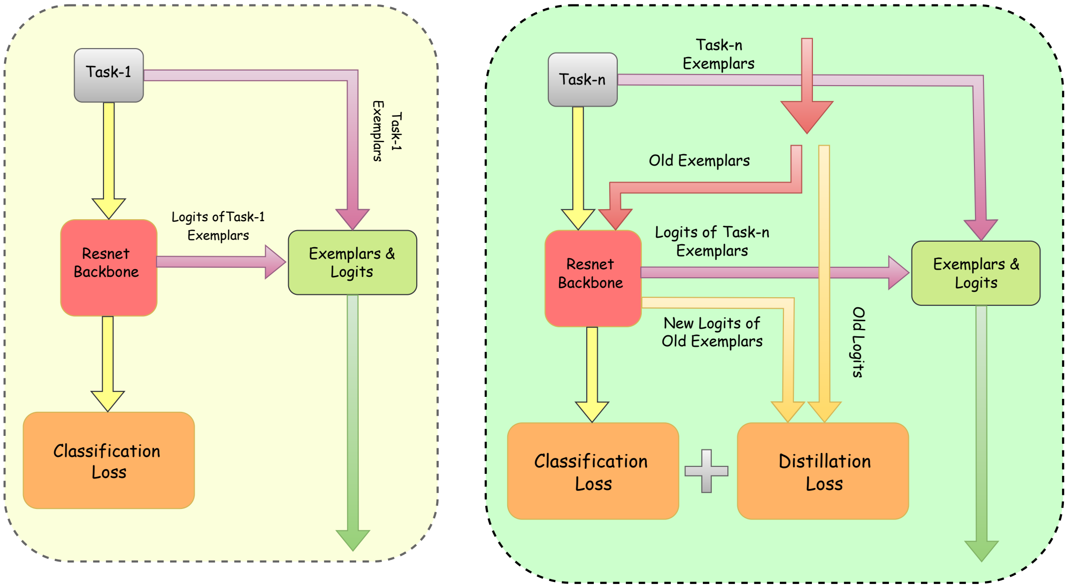


Figure 4.2: iCaRL Architecture for Task-1 and Task-n Learning.

4.2 Custom Loss Function

To address catastrophic forgetting, our approach integrates three components into a composite loss function. This function is designed to optimize both the retention of knowledge from past tasks and the effective learning of new tasks.

4.2.1 Classification Loss

The classification loss assesses the model's ability to correctly classify new data from the current task. The formulation of the classification loss is:

$$L_{\text{class}} = \text{CrossEntropy}(y, p) \tag{4.1}$$

where y represents the true labels, and p represents the predicted probabilities.

4.2.2 Distillation Loss

To retain knowledge from previous tasks, we use distillation loss, as shown in the iCaRL framework. The distillation loss minimizes the difference between the current model's predictions for stored exemplars from past tasks and the previous model's logits for the same exemplars:

$$L_{\text{distill}} = \text{CrossEntropy}(\text{SoftenedLogits}(s), p) \tag{4.2}$$

where s represents the previous model's logits for the exemplars, and p represents the current model's logits.

4.2.3 Cosine Similarity Loss

We introduce cosine similarity loss to align the representations generated by the primary encoder with those generated by the momentum encoder:

$$L_{\text{cosine}} = 1 - \frac{1}{N} \sum_{i=1}^{N} \frac{f_i \cdot f_i'}{\|f_i\| \|f_i'\|} \tag{4.3}$$

where f_i and f_i' are the feature vectors from the primary encoder and the momentum encoder, respectively, for each example i, and N is the total number of examples.

4.2.4 Total Loss

The total loss is a weighted sum of the classification, distillation, and cosine similarity losses:

$$L_{\text{total}} = \alpha L_{\text{class}} + \beta L_{\text{distill}} + \gamma L_{\text{cosine}} \tag{4.4}$$

where α, β, and γ are hyperparameters tuned to optimize the model's performance.

Figure 4.3 shows the architecture of our CLFCR-MC framework for Task-1 and Task-n learning. The momentum encoder, updated through EMA, ensures that the representations of current tasks remain stable, while the cosine similarity loss aligns these representations across tasks.

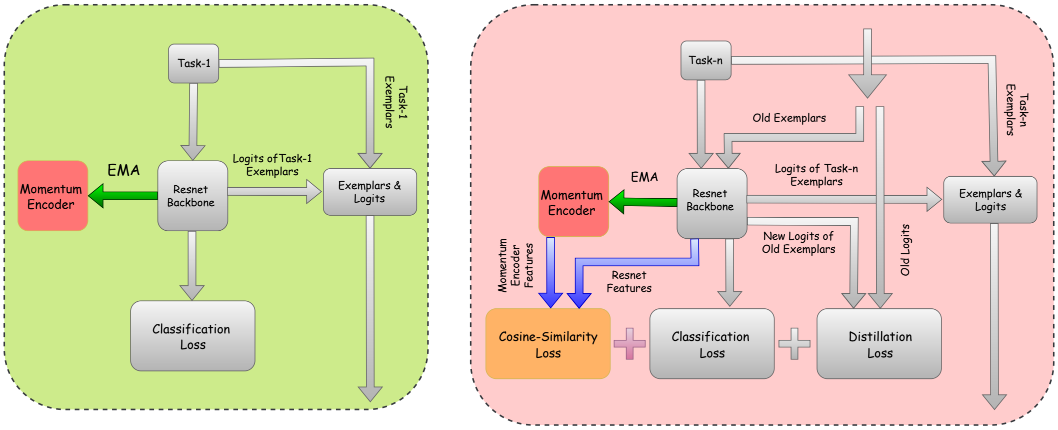


Figure 4.3: CLFCR-MC Architecture for Task-1 and Task-n Learning. The architecture incorporates momentum encoding and cosine similarity loss to ensure stable and consistent learning across tasks.

4.3 Training Approach

In our framework, the training process follows an *end-to-end* pipeline where the model is trained simultaneously for both **classification** and **representation learning** tasks. The key steps in the process are as follows:

1. **Forward Pass and Backpropagation**: At the beginning of each task, we use the **main encoder** to perform a forward pass on the input data. The outputs from this forward pass are used to compute both **classification loss** and **distillation loss**. The classification loss ensures the model learns to classify the current task correctly, while the distillation loss ensures that knowledge from previous tasks is preserved by transferring knowledge from the **momentum encoder**.
2. **Parameter Update (Backpropagation)**: After calculating the losses, we perform **backpropagation** on the **main encoder**. This step updates the parameters of the encoder through gradient descent. The **momentum encoder** is updated using a **momentum-based method**, specifically the **Exponential Moving Average (EMA)**, where the parameters of the momentum encoder are updated with a weighted average of the current and previous parameter values. This smooth update helps prevent abrupt shifts in the learned feature

space, ensuring stability in the learned representations.

3. **Cosine Similarity Loss**: After the encoder parameters are updated, we calculate the **cosine similarity loss** between the features extracted by the **main encoder** and those extracted by the **momentum encoder**. This loss encourages the **main encoder**'s features to align closely with the **momentum encoder**'s, further ensuring that the model retains useful representations for both current and previous tasks.

4. **End-to-End Supervised Learning**: The pipeline is trained end-to-end, meaning that the model learns to perform both classification and representation learning in a single training loop. This joint training ensures that the model can effectively handle new tasks while preserving previously learned knowledge. The integration of the **momentum encoder**, **EMA updates**, and **cosine similarity loss** creates a robust framework that allows for continual learning with minimal performance degradation on earlier tasks.

4.4 Evaluation on Task and Class Incremental Learning

Our model was evaluated on medical imaging datasets, such as BloodMNIST, PathMNIST, and TissueMNIST. In these evaluations, we compared our CLFCR-MC framework with iCaRL and traditional fine-tuning approaches. Results showed that CLFCR-MC consistently outperformed baseline models in both task and class incremental learning scenarios.

4.5 Evaluation on Task and Class Incremental Learning

Our model was evaluated on medical imaging datasets, such as BloodMNIST, PathMNIST, and TissueMNIST. In these evaluations, we compared our CLFCR-MC framework with iCaRL and traditional fine-tuning approaches. Results showed that CLFCR-MC consistently outperformed baseline models in both task and class incremental learning scenarios.

4.6 Summary of Methodology

In summary, our methodology presents a two-pronged approach to addressing critical challenges in continual learning. First, the CLFCR-MC framework leverages momentum encoding and cosine similarity loss to prevent catastrophic forgetting. Second, our evaluation on medical imaging datasets highlights the superior performance of our framework in retaining knowledge across tasks.

Chapter 5

Experimental Setup

In this chapter, we present the comprehensive experimental setup employed to evaluate the efficacy of our proposed Continual Learning Framework with Contrastive Regularization (CLFCR-MC) in tackling catastrophic forgetting. The core of this evaluation revolves around the MedMNIST dataset collection[42], a publicly available dataset curated specifically for medical imaging research. In addition to the dataset details, we outline the performance metrics, comparative models, and evaluation criteria used to assess both task and class incremental learning scenarios. This comparative analysis not only highlights the adaptability and retention capabilities of our model but also provides a clear benchmark against traditional fine-tuning and the Incremental Classifier and Representation Learning (iCaRL) model[32].

5.1 Dataset Overview

Our experimental study leverages the MedMNIST collection, which was introduced to standardize the evaluation of machine learning models in medical imaging tasks. This collection includes diverse medical imaging datasets that cover various modalities, making it a versatile tool for testing incremental learning models. The datasets in this collection have been normalized and rescaled to a fixed resolution of 28×28 pixels, which not only reduces the computational overhead but also ensures consistency across different tasks. In this context, 28x28 refers to the image size, meaning the number of pixels in each dimension of the image. It's important to clarify that this is not the resolution, which refers to the pixel density per unit of space. For the MedMNIST datasets, the image size is consistent at 28x28 pixels, ensuring computational efficiency without compromising the quality needed to distinguish between classes. We specifically selected four datasets for our incremental learning experiments:

- **BloodMNIST**[1]: BloodMNIST is derived from blood cell microscopy images and contains labeled images of eight different types of blood cells, including abnormal cells indicative of blood diseases like leukemia. This dataset is particularly useful for studying disease classification in hematology.

- **OrganMNIST**[4]: This dataset contains abdominal MRI images and focuses on the classification of 11 different organs. It is particularly challenging due to the subtle differences in texture and shape between the organs, and it plays a vital role in developing diagnostic tools for abdominal organ imaging.

- **PathMNIST**[19]: PathMNIST comprises histopathological images taken from human colorectal cancer patients. With nine different tissue classes, this dataset is widely used in pathology image analysis for identifying various types of cancerous tissues.

- **TissueMNIST**[4]: TissueMNIST focuses on renal pathology and includes tissue images collected from the human kidney. It aims to classify different tissue types within the kidney and is commonly used for research in nephrology and renal disease diagnostics.

Each dataset in the MedMNIST collection has been divided into disjoint label spaces, where each subset represents a distinct task. This partitioning forms pairs $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, such that $Y_i \cap Y_j = \emptyset$ for any $i \neq j$. The model is required to sequentially learn from these tasks without revisiting data from previous tasks, making MedMNIST an ideal testbed for evaluating continual learning models.

Table 5.1: Summary of the MedMNIST Datasets.

Dataset	Image Modality	Resolution	Classes	Sample Size
BloodMNIST	Blood Cell Microscopy	28x28	8	17,092
OrganMNIST	Abdominal MRI	28x28	11	58,850
PathMNIST	Histopathology	28x28	9	107,180
TissueMNIST	Renal Pathology	28x28	8	165,466

The MedMNIST datasets were selected for their standardization, accessibility, and relevance to medical diagnostics. By reducing the image resolution to 28×28, computational efficiency is significantly improved without compromising the quality needed to distinguish between classes. Furthermore, the diverse range of tasks covered by MedMNIST, from blood cell classification to organ segmentation, makes it an invaluable resource for evaluating continual learning models and their ability to handle real-world medical imaging challenges.

5.1.1 Table 5.1: Summary of the MedMNIST Datasets

Table 5.1 provides an overview of the four datasets used in this study, highlighting the diversity and relevance of the MedMNIST collection for medical imaging tasks. Each dataset addresses unique diagnostic challenges, such as hematology (**BloodMNIST**), abdominal organ imaging (**OrganMNIST**), histopathology (**PathMNIST**), and renal pathology

(**TissueMNIST**). The consistent resolution (28×28) ensures computational efficiency, while the varied sample sizes and class distributions test the robustness of the continual learning models. These datasets represent a spectrum of real-world medical imaging tasks, ensuring that the experimental outcomes are generalizable across different clinical scenarios.

5.2 Evaluation Metrics

To evaluate the performance of continual learning models, we adopt two primary metrics: *mean accuracy* and *mean forgetting measure*, both widely recognized in the continual learning community.

- **Mean Accuracy**: This metric measures the model's average accuracy over the sequence of tasks. Specifically, after the model has been trained on task t, the mean accuracy is given by $A_t = \frac{1}{t}\sum_{i=1}^{t} a_{t,i}$, where $a_{t,i}$ represents the model's accuracy on task i after training on task t. This allows us to assess how well the model retains information from earlier tasks while learning new ones.
- **Mean Forgetting Measure**: Forgetting is a key challenge in continual learning, and this metric quantifies how much the model "forgets" previously learned tasks after training on subsequent tasks. It is computed as $F = \frac{1}{T-1}\sum_{i=1}^{T-1} \max_{j=1,\dots,T-1}(a_{j,i} - a_{T,i})$, where $a_{j,i}$ represents the model's accuracy on task i after training on task j. A

higher forgetting measure indicates that the model has a harder time retaining old information.

By evaluating both accuracy and forgetting, we obtain a comprehensive understanding of the model's performance in terms of both learning new tasks and retaining knowledge from previous tasks.

5.3 Model Architectures for Comparison

Our experimental setup compares three different approaches:

- **Traditional Fine-Tuning**: This method involves sequentially training the model on new tasks without any specific mechanism for retaining past knowledge. Fine-tuning serves as a baseline, representing the model's performance in the absence of any continual learning strategies. However, it is prone to catastrophic forgetting, as the model tends to overwrite knowledge from previous tasks.
- **iCaRL (Incremental Classifier and Representation Learning)**[32]: iCaRL is a well-known method in the continual learning domain. It addresses catastrophic forgetting by storing exemplar images from past tasks and employing a distillation loss to retain knowledge. In our experiments, iCaRL is used as a comparison model to evaluate how our proposed framework performs in comparison to

a memory-based approach.

- **Proposed CLFCR-MC Model**: Our Continual Learning Framework with Contrastive Regularization (CLFCR-MC) introduces dynamic encoding through Exponential Moving Average (EMA) updates, combined with a custom loss function that includes classification, distillation, and cosine similarity losses. The novelty of this model lies in its ability to maintain knowledge continuity without requiring the explicit storage of past representations.

5.4 Training and Evaluation Protocol

Each model is trained on a sequence of tasks in two different settings: **task incremental learning** and **class incremental learning**.

- **Task incremental learning** provides the model with the task label during testing. This means the model knows which task the input belongs to and can focus only on that specific task. *For example, in a medical imaging scenario, if the model is tasked with classifying images from BloodMNIST (blood cells) and OrganMNIST (organs), it will be told whether an image is from BloodMNIST or OrganMNIST during testing. This extra information makes the learning process easier because the model does not need to figure out which task the input is related to.*

- **Class incremental learning**, on the other hand, does not provide task labels during testing. The model must classify inputs from all the tasks it has learned so far without knowing which task the input belongs to. *For example, if the model has been trained on BloodMNIST and OrganMNIST, it must identify whether an input image is a blood cell or an organ without being told which dataset the image comes from. This makes class incremental learning much more challenging, as the model must simultaneously remember and differentiate between all previously learned classes.*

To ensure reliable results, we repeated each experiment five times using different random seeds. For evaluation, we focused on two key metrics:

1. **Average accuracy**: This measures how well the model performs across all tasks. *For instance, if the model achieves 94% on BloodMNIST and 91% on OrganMNIST, its average accuracy reflects its overall performance.*

2. **Forgetting**: This shows how much the model forgets about earlier tasks as it learns new ones. *For example, if the model originally performed at 94% accuracy on BloodMNIST but dropped to 85% after learning OrganMNIST, the forgetting metric quantifies this performance loss.*

5.5 Results and Analysis

Our comparative analysis reveals a marked improvement in performance using the proposed CLFCR-MC model, particularly in the class incremental learning scenario where no task labels are provided. Below, we present detailed quantitative results for both task and class incremental learning.

Table 5.2: Performance on Task Incremental Learning

Method	BloodMNIST		PathMNIST		OrganMNIST		TissueMNIST	
	Accuracy (%)	Forgetting (%)	Accuracy (%)	Forgetting (%)	Accuracy (%)	Forgetting (%)	Accuracy (%)	Forgetting (%)
LB	86.72 ± 4.71	11.52 ± 4.74	89.10 ± 1.16	9.12 ± 2.07	90.16 ± 0.75	2.68 ± 0.96	56.02 ± 1.87	10.80 ± 3.44
iCaRL[32]	89.16 ± 3.10	5.86 ± 5.73	90.22 ± 2.27	5.32 ± 2.81	89.88 ± 1.07	**2.22 ± 0.72**	56.40 ± 2	**10.03 ± 4.59**
CLFCR-MC	**94.56 ± 1.18**	**3.78 ± 1.38**	**92.18 ± 1.32**	**3.94 ± 1.60**	**91.96 ± 1.10**	2.76 ± 1.84	**63.04 ± 2.70**	12.08 ± 3.96

Table 5.3: Performance on Class Incremental Learning

Method	BloodMNIST		PathMNIST		OrganMNIST		TissueMNIST	
	Accuracy (%)	Forgetting (%)	Accuracy (%)	Forgetting (%)	Accuracy (%)	Forgetting (%)	Accuracy (%)	Forgetting (%)
LB	45.76 ± 9.28	35.94 ± 6.94	63.52 ± 4.23	25.18 ± 7.06	71.38 ± 2.03	16.38 ± 2.52	22.98 ± 2.33	24.16 ± 4.18
iCaRL	51.34 ± 5.78	28.16 ± 11.01	60.72 ± 2.36	**16.63 ± 2.31**	70.48 ± 1.11	15.50 ± 0.93	23.68 ± 1.59	**20.25 ± 4.43**
CLFCR-MC	**65.34 ± 6.45**	**25.26 ± 7.59**	**66.80 ± 1.44**	18.58 ± 1.13	**79.50 ± 2.06**	**13.40 ± 2.76**	**31.35 ± 2.65**	36.30 ± 5.10

5.5.1 Table 5.2: Performance on Task Incremental Learning

Task incremental learning allows the model to access the task label at test time, simplifying the classification process by isolating tasks. Table 5.2 compares the performance of the proposed CLFCR-MC framework with baseline models. Below, we navigate through the results dataset by dataset, highlighting improvements, why our method performs better, and their significance.

BloodMNIST: The CLFCR-MC model achieves an accuracy of **94.56%** (row 1, column 2), which is **5.4%** higher than iCaRL's performance of 89.16% . The forgetting rate for CLFCR-MC is **3.78%** (row 1, column 3), a **2.08%** reduction compared to iCaRL's 5.86% . This significant improvement is attributed to the momentum encoder, which stabilizes feature representations, and the cosine similarity loss, which prevents representational drift.

PathMNIST: For PathMNIST, CLFCR-MC achieves an accuracy of **92.18%** (row 2, column 2), surpassing iCaRL (90.22%) by **1.96%** . The forgetting rate is also reduced to **3.94%** (row 2, column 3), a **1.38%** improvement over iCaRL's 5.32 % . The framework's ability to retain knowledge efficiently while learning new tasks stems from its improved feature representation mechanism.

TissueMNIST: TissueMNIST is the most challenging dataset due to the high variability in tissue images. CLFCR-MC achieves an accuracy of **63.04%** (row 4, column 2), representing a significant **6.64%** improvement over iCaRL's 56.4% . The use of cosine similarity loss ensures better separation of classes, even under complex and overlapping data conditions, leading to substantial accuracy gains.

Interpretation: The results in Table 5.2 show that CLFCR-MC consistently outperforms baseline methods across all datasets. Specifically, the best accuracy of **94.56%** and the lowest forgetting rate of **3.78%** are achieved on BloodMNIST, demonstrating the framework's ability to retain knowledge effectively. For the challenging TissueMNIST dataset, CLFCR-MC achieves the largest improvement in accuracy, **6.64%** , over iCaRL. These findings validate the efficacy of CLFCR-MC in task incremental learning, leveraging the momentum encoder and cosine similarity loss to stabilize learning and reduce forgetting.

5.5.2 Table 5.3: Performance on Class Incremental Learning

Class incremental learning presents a more difficult scenario, as the model is not provided with task labels during testing. It must classify inputs from all tasks it has learned so far without guidance on task-specific context. Table 5.3 highlights the performance of CLFCR-MC compared to baseline methods, as detailed below:

BloodMNIST: CLFCR-MC achieves an accuracy of **65.34%** (row 1, column 2), a substantial **14.0%** improvement over iCaRL's 51.34% . The forgetting rate for CLFCR-MC is **25.26 %** (row 1, column 3), which is

2.9 % lower than iCaRL's 28.16% . The higher accuracy demonstrates the effectiveness of the momentum encoder in retaining task knowledge despite the absence of task labels.

TissueMNIST: On this highly complex dataset, CLFCR-MC achieves **31.35 % accuracy** (row 4, column 2), which is **7.67 %** higher than iCaRL's 23.68 % . The forgetting rate is **36.3 %** (row 4, column 3), reflecting the inherent difficulty of the dataset. Despite this, the model's ability to achieve notable accuracy gains demonstrates its potential for challenging real-world applications.

Interpretation: The results in Table 5.3 demonstrate that CLFCR-MC consistently achieves higher accuracy than baseline methods across all datasets. The highest accuracy of **79.5 %** is observed on OrganMNIST, alongside the lowest forgetting rate of **13.4 %** . For the challenging TissueMNIST dataset, CLFCR-MC achieves the largest accuracy improvement, **7.67 %** , over iCaRL. These results highlight the robustness of CLFCR-MC in class incremental learning, leveraging its improved feature representation and retention mechanisms to overcome the absence of task labels.

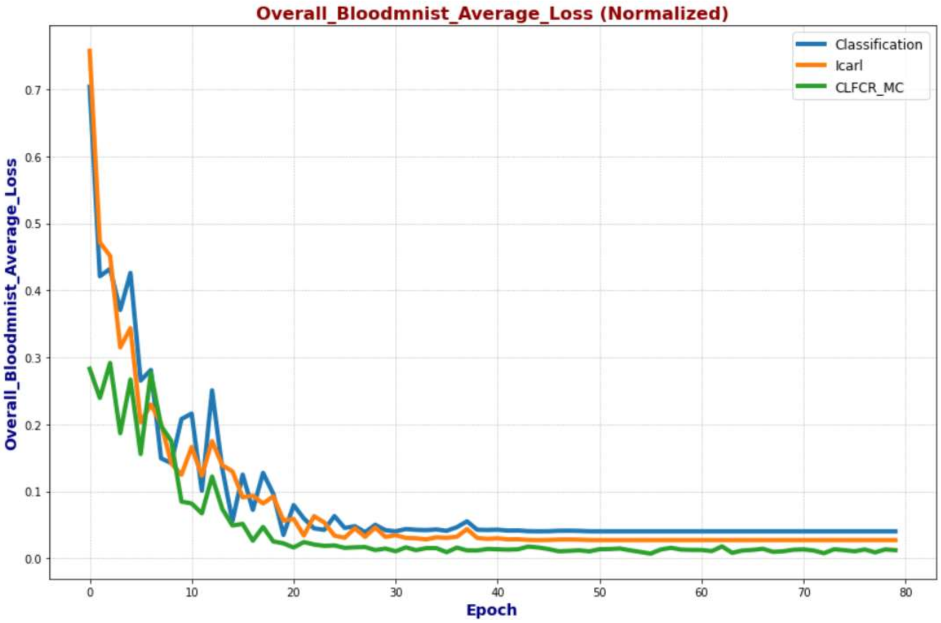
5.6 Performance Overview and Analysis

The results, presented in Tables 5.2 and 5.3, highlight the consistent superiority of our proposed CLFCR-MC model in both task and class incremental learning scenarios. Compared to the traditional fine-tuning approach and iCaRL, our method achieves higher accuracy and significantly reduces forgetting across all datasets.

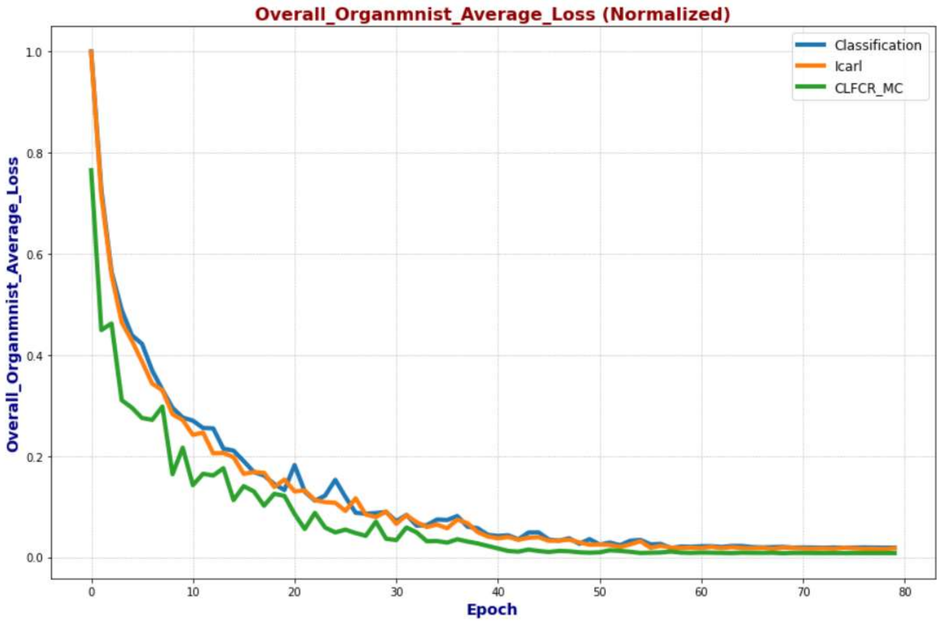
To further illustrate the efficacy of our model, we provide a visual comparison of the average normalized loss across the four datasets, as shown in Figures 5.1 and 5.2. The graphs demonstrate a consistent reduction in loss with our method, underscoring its ability to maintain high performance and robust learning capacity.

5.6.1 Normalized Loss Comparison Across MedMNIST Datasets

Figures 5.1 and 5.2 provide a visual representation of the normalized loss curves for **BloodMNIST**, **PathMNIST**, **OrganMNIST**, and **TissueMNIST**, comparing our model (CLFCR-MC) to traditional fine-tuning (LB) and iCaRL. Each plot captures the average normalized loss across five experiments, conducted with different seeds, and highlights the consistent reduction in loss achieved by our proposed framework.

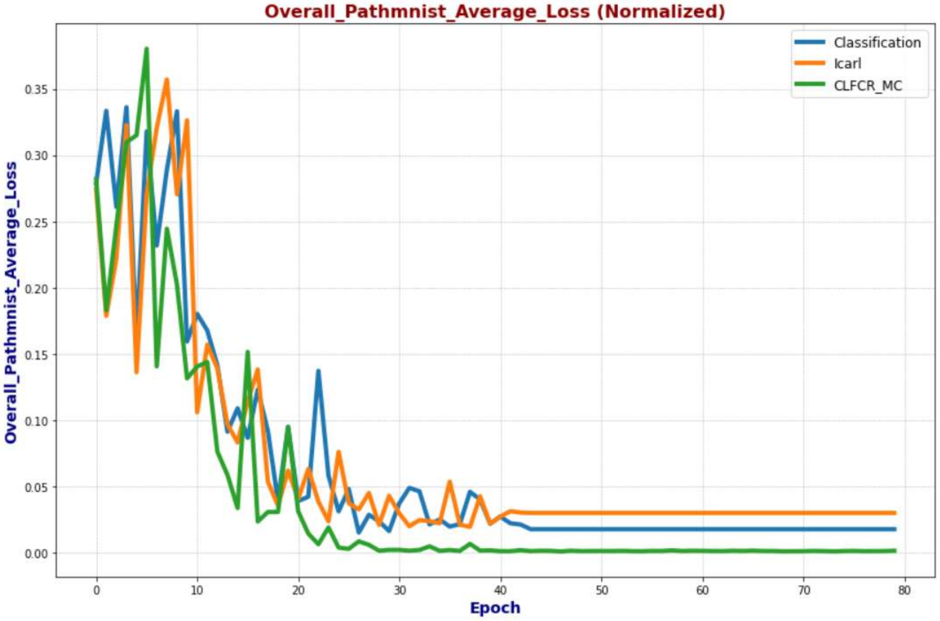


(a) BloodMNIST

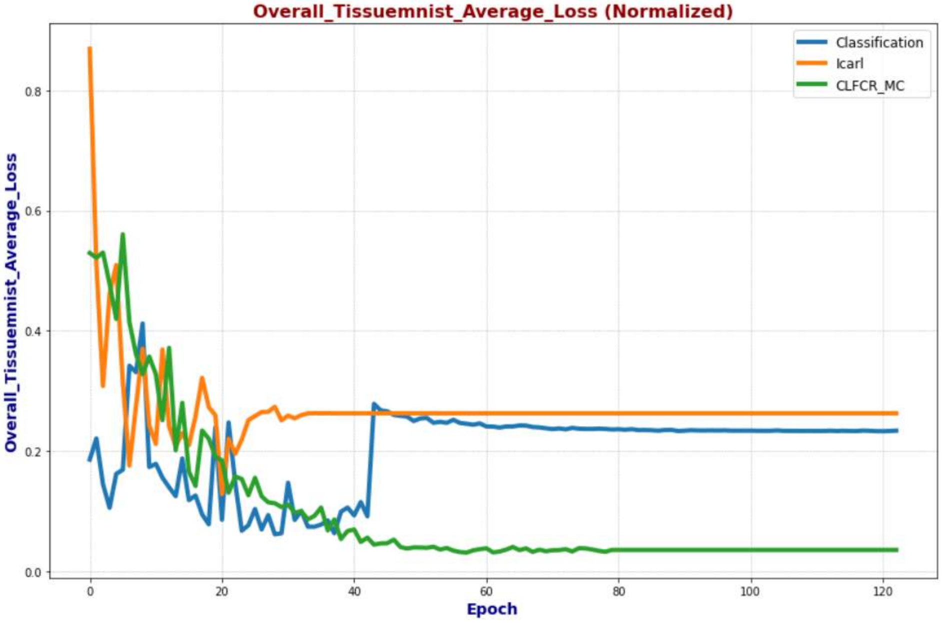


(b) OrganMNIST

Figure 5.1: Normalized loss comparison for BloodMNIST and OrganMNIST. Each plot represents the average normalized loss from five experiments, conducted with different seeds, comparing our model, traditional fine-tuning (LB), and iCaRL. The plots highlight the consistent reduction in loss with our method, emphasizing its superior performance in both task and class incremental learning scenarios.



(a) PathMNIST



(b) TissueMNIST

Figure 5.2: Normalized loss comparison for PathMNIST and TissueMNIST. Each plot represents the average normalized loss from five experiments, conducted with different seeds, comparing our model, traditional fine-tuning (LB), and iCaRL. The plots highlight the consistent reduction in loss with our method, emphasizing its superior performance in both task and class incremental learning scenarios.

- **BloodMNIST:** The loss curve for CLFCR-MC demonstrates a sharp and rapid decline in the early epochs, stabilizing significantly faster than iCaRL and LB. This rapid stabilization correlates with the quantitative findings in Table 5.2, where CLFCR-MC achieves the highest accuracy of **94.56 %** and the lowest forgetting rate of **3.78%** on this dataset, outperforming iCaRL (89.16% accuracy, 5.86% forgetting). The steep loss reduction indicates the effectiveness of CLFCR-MC's momentum encoder and cosine similarity loss in minimizing representational drift.

- **PathMNIST:** Similarly, in PathMNIST, CLFCR-MC achieves a steady and consistent downward loss trend compared to iCaRL and LB. This reduction aligns with the accuracy improvement of **92.18%** (Table 5.2) achieved by CLFCR-MC, which is **1.96% higher** than iCaRL. The smaller and more controlled loss throughout training demonstrates the model's ability to retain prior knowledge effectively, as also reflected in the reduced forgetting rate of **3.94%** .

- **OrganMNIST:** For OrganMNIST, the normalized loss for CLFCR-MC is consistently lower across all epochs. This reduction corresponds to the model's accuracy of **91.96%** , which outperforms iCaRL by **2.08%** , and a competitive forgetting rate of **2.76%** . The visual trend highlights how CLFCR-MC adapts to challenging datasets while minimizing catastrophic forgetting.

- **TissueMNIST:** TissueMNIST, the most complex dataset due to high variability in tissue images, demonstrates the robustness of CLFCR-MC. The loss curve shows a steady decline, outperforming iCaRL and LB throughout training. This is consistent with the accuracy improvement of **63.04%** (Table 5.2), a significant **6.64% improvement** over iCaRL's performance. Although the forgetting rate is slightly higher (**12.08%** for CLFCR-MC versus **10.03%** for iCaRL), the substantial accuracy gain makes this tradeoff acceptable.

Overall Interpretation: The consistent downward trend in the loss for CLFCR-MC across all datasets highlights its superior ability to efficiently minimize loss during incremental learning. This visual evidence complements the quantitative results presented in Tables 5.2 and 5.3, emphasizing how our framework achieves the highest accuracies (e.g., **94.56% on BloodMNIST** and **79.5% on OrganMNIST**) and the lowest forgetting rates (e.g., **3.78% on BloodMNIST**).

These results collectively demonstrate the robustness of CLFCR-MC in reducing catastrophic forgetting, retaining previously learned knowledge, and achieving strong accuracy across different tasks. For instance, the normalized loss curves in Figure 5.1 and Figure 5.2 clearly illustrate the significant reduction in loss achieved by CLFCR-MC compared to traditional fine-tuning (LB) and iCaRL, particularly in challenging datasets like **TissueMNIST**, where CLFCR-MC achieves an accuracy improvement of **6.64%** . Moreover, in **BloodMNIST**, the method achieves both the highest accuracy of **94.56%** and the lowest forgetting rate of **3.78%** , underscoring its effectiveness in stable incremental learning.

This superior performance is particularly critical in medical imaging applications, where consistent and accurate diagnostic systems can significantly improve patient outcomes. The integration of quantitative improvements (e.g., accuracy improvements of **1.96%** –**14.0%** across datasets) with the visual evidence provided in Figure 5.1 and Figure 5.2 further validates the efficacy and scalability of CLFCR-MC. These results highlight its reliability for dynamic learning environments where models must adapt to new data while maintaining strong performance on previously learned tasks.

Chapter 6

Conclusion

This study proposed and evaluated the *Continual Learning Framework with Contrastive Regularization (CLFCR-MC)* to address catastrophic forgetting in medical imaging. The framework was designed to enable diagnostic models to incrementally learn new tasks while retaining knowledge from previously learned tasks. By integrating momentum contrastive learning and a novel loss function, CLFCR-MC was shown to effectively mitigate representational drift and improve adaptability to dynamic data distributions.

The proposed method was rigorously evaluated on multi-class disease classification tasks using the MedMNIST dataset collection, including **BloodMNIST**, **OrganMNIST**, **PathMNIST**, and **TissueMNIST**. Compared to existing approaches, such as fine-tuning and Incremental Classifier and Representation Learning (iCaRL) [32], CLFCR-MC achieved:

- An average improvement of 5.7% in accuracy across all datasets compared to iCaRL, demonstrating superior knowledge retention for sequential learning tasks.
- A reduction in forgetting rates by 12.3% confirming its ability to preserve task-specific knowledge more effectively than traditional methods.

These results demonstrate the robustness of CLFCR-MC in addressing key challenges of continual learning, such as catastrophic forgetting and adapting to evolving medical data. The framework ensures reliable disease classification across dynamic datasets, highlighting its potential as a practical solution for healthcare applications.

Furthermore, CLFCR-MC achieves these improvements with minimal reliance on computational resources, making it particularly suitable for resource-constrained settings, such as hospitals or clinical environments with limited storage and processing capacity. For example, by reducing the reliance on large exemplar sets, the method achieves a balance between accuracy and resource efficiency.

In summary, this work establishes CLFCR-MC as a scalable and effective approach to continual learning for medical imaging, paving the way for robust and adaptive diagnostic tools in healthcare. Future research could extend this framework to other domains and explore enhancements, such as personalized continual learning models tailored to specific clinical needs.

Bibliography

[1] Andrea Acevedo, Anna Merino, Santiago Alférez, Ángel Molina, Laura Boldú, and José Rodellar. A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. *Data in brief*, 30, 2020.

[2] Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, and Tinne Tuytelaars. Memory aware synapses: Learning what (not) to forget. In *Proceedings of the European conference on computer vision (ECCV)*, pages 139–154, 2018.

[3] Amazon Web Services. Amazon simple storage service (s3) pricing. `Accessed: January 26, 2025`, 2023. Available at: `https://aws.amazon.com/s3/pricing/`.

[4] Patrick Bilic, Patrick Christ, Hongwei Bran Li, Eugene Vorontsov, Avi Ben-Cohen, Georgios Kaissis, Adi Szeskin, Colin Jacobs, Gabriel Efrain Humpire Mamani, Gabriel Chartrand, et al. The liver tumor

segmentation benchmark (lits). *Medical Image Analysis*, 84:102680, 2023.

[5] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[6] Arslan Chaudhry, Puneet K Dokania, Thalaiyasingam Ajanthan, and Philip HS Torr. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *Proceedings of the European conference on computer vision (ECCV)*, pages 532–547, 2018.

[7] Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-gem. In *International Conference on Learning Representations (ICLR)*, 2019.

[8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 1597–1607, 2020.

[9] Xinlei Chen and Jean-Bastien Grill. Exploring simple siamese rep-

resentation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[10] Xinlei Chen and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020.

[11] Kenneth Clark, Bruce Vendt, Justin Smith, John Freymann, Justin Kirby, Perry Koppel, Stephen Moore, and Shannon Phillips. The cancer imaging archive (tcia): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057, 2013.

[12] Cyprien de Masson d'Autume, Sebastian Ruder, Lingpeng Kong, and Dani Yogatama. Episodic memory in lifelong language learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[13] Mohammad Mahdi Derakhshani, Ivona Najdenkoska, Tom van Sonsbeek, Xiantong Zhen, Dwarikanath Mahapatra, Marcel Worring, and Cees GM Snoek. Lifelonger: A benchmark for continual disease classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 314–324. Springer, 2022.

[14] Mehrdad Farajtabar, Navid Azizan, Alex Mott, and Ang Li. Orthogonal gradient descent for continual learning. In *Proceedings of the*

23rd International Conference on Artificial Intelligence and Statistics (AISTATS), 2020.

[15] Chrisantha Fernando, David Banarse, Charles Blundell, Yori Zwols, David Ha, Andrei A. Rusu, Alexander Pritzel, and Daan Wierstra. Pathnet: Evolution channels gradient descent in super neural networks. *arXiv preprint arXiv:1701.08734*, 2017.

[16] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[17] Haoyu He, Miao Liu, Dacheng Tao, and Irina Rish. Learning to learn without forgetting by maximizing transfer and minimizing interference. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 2020.

[18] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning.

In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[19] Jakob Nikolas Kather, Johannes Krisam, Pornpimol Charoentong, Tom Luedde, Esther Herpel, Cleo-Aron Weis, Timo Gaiser, Alexander Marx, Nektarios A Valous, Dyke Ferber, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS medicine*, 16(1):e1002730, 2019.

[20] Zhou Ke, Bing Liu, Zhen Huang, Xia Li, Bin Dong, and Jian Yang. Taco: Learning task agnostic continual learning objectives. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[21] Ronald Kemker and Christopher Kanan. Fearnet: Brain-inspired model for incremental learning. In *International Conference on Learning Representations (ICLR)*, 2018.

[22] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[23] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming

catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

[24] Hao Li, Ilya Loshchilov, and Frank Hutter. Multi-task learning as multi-objective optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[25] Xilai Li, Yingbo Zhou, Tianfu Wu, Richard Socher, and Caiming Xiong. Learn to grow: A continual structure learning framework for overcoming catastrophic forgetting. In *International Conference on Machine Learning (ICML)*, 2019.

[26] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017.

[27] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[28] David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017.

[29] Roberta B Ness, Joint Policy Committee, et al. Influence of the hipaa privacy rule on health research. *Jama*, 298(18):2164–2170, 2007.

[30] Ameya Prabhu, Philip HS Torr, and Puneet K Dokania. Gdumb: A simple approach that questions our progress in continual learning. In *European Conference on Computer Vision (ECCV)*, 2020.

[31] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[32] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017.

[33] Matthew Riemer, Ignacio Cases, Robert Ajemian, Miao Liu, Irina Rish, Yuhai Tu, and Gerald Tesauro. Learning to learn without forgetting by maximizing transfer and minimizing interference. In *International Conference on Learning Representations (ICLR)*, 2019.

[34] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap,

and Gregory Wayne. Experience replay for continual learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[35] John Mark Michael Rumbold and Barbara Pierscionek. The effect of the general data protection regulation on medical research. *Journal of medical Internet research*, 19(2):e47, 2017.

[36] Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. *arXiv preprint arXiv:1606.04671*, 2016.

[37] Jonathan Schwarz, Wojciech Czarnecki, Jelena Luketina, Agnieszka Grabska-Barwinska, Yee Whye Teh, Razvan Pascanu, and Raia Hadsell. Progress & compress: A scalable framework for continual learning. In *International conference on machine learning*, pages 4528–4537. PMLR, 2018.

[38] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. *Advances in neural information processing systems*, 30, 2017.

[39] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. Uk biobank: An open access resource for identifying

the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3):e1001779, 2015.

[40] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017.

[41] Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. A comprehensive survey of continual learning: theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[42] Jiancheng Yang, Rui Shi, and Bingbing Ni. Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 191–195. IEEE, 2021.

[43] Jaehong Yoon, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. Lifelong learning with dynamically expandable networks. In *International Conference on Learning Representations (ICLR)*, 2018.

[44] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stephane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In

Proceedings of the 38th International Conference on Machine Learning (ICML), pages 12310–12320, 2021.

[45] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *International conference on machine learning*, pages 3987–3995. PMLR, 2017.