Continual learning for energy management systems: A review of methods and applications, and a case study

Aya Nabil Sayed, Yassine Himeur, Iraklis Varlamis, Faycal Bensaali

Applied Energy 384 (2025) 125458

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Continual learning for energy management systems: A review of methods and applications, and a case study

Aya Nabil Sayed [a,*], Yassine Himeur [b], Iraklis Varlamis [c], Faycal Bensaali [a]

[a] Department of Electrical Engineering, College of Engineering, Qatar University, Doha, Qatar
[b] College of Engineering and Information Technology, University of Dubai, Dubai, United Arab Emirates
[c] Department of Informatics and Telematics, Harokopio University of Athens, Athens, Greece

ARTICLE INFO

Keywords:
Continual learning
Lifelong learning
Deep learning
Catastrophic forgetting
Energy management systems
Non-intrusive load monitoring
Demand-side management
Fault/anomaly detection
Load forecasting
Renewable energy integration

ABSTRACT

An intelligent system must incrementally acquire, update, accumulate, and exploit knowledge to navigate the real world's intricacies. This trait is frequently referred to as Continual Learning (CL), and it can be limited by catastrophic forgetting, a phenomenon in which learning a new task acutely reduces the system's performance on prior tasks. Numerous strategies have been developed to address this issue, as CL is essential for developing Artificial Intelligence (AI) systems that adapt to dynamic environments. This study examines the practical applications of CL, concentrating on energy management systems and their integration with Deep Learning (DL) models. Energy management systems are strategies and methods for monitoring, controlling, and optimizing energy use within a system or organization. The literature is systematically analyzed, highlighting methods such as replay techniques, regularization strategies, and architectural adaptations that address the challenges of catastrophic forgetting. Moreover, the review encompasses various energy-related applications, including non-intrusive load monitoring, demand-side management, fault/anomaly detection, load forecasting/prediction, and renewable energy integration. Additionally, a case study on anomaly detection in energy systems is conducted, comparing different CL approaches. The case study findings aim to bridge the gap between theoretical advancements and real-world applications, providing insights and guidelines for implementing CL in diverse fields. Finally, this survey identifies key challenges that impede the deployment of CL and suggests potential directions to enhance its implementation in the energy management sector.

1. Introduction

With climate change posing ever-increasing challenges, global leaders and governments must take aggressive decarbonization activities to ensure societal sustainability [1,2]. Building energy systems provide tenant comfort and meet necessities, accounting for around 34% of world energy consumption and 37% of global greenhouse gas emissions [3,4]. Deep Learning (DL)-powered modeling approaches have recently been applied to many elements of building energy systems to increase energy efficiency, including energy prediction [5,6], predictive maintenance [7,8], and control optimization [9]. The continuous development of Machine Learning (ML)/DL techniques has the potential to advance innovative management and accelerate the decarbonization of building energy systems. Overall, ML/DL models have demonstrated their effectiveness by rivaling or outperforming human performance on various energy-related tasks [10,11]. While these achievements are impressive, they were acquired using static models incapable of responding to changing conditions over time. This limitation means the training procedure must be restarted whenever new data is available. In today's dynamic environment, this method soon becomes unattainable for continuous data streams owing to storage restrictions or privacy concerns [12]. This underscores the urgent and immediate need for systems that can constantly adapt and learn over time to keep pace with the evolving data landscape [13]. In Continual Learning (CL), tasks are learned sequentially but treated as if they were learned simultaneously.

CL (also known as lifelong learning [14], Incremental Learning (IL) [15], or sequential learning [16]) is a paradigm within ML rather than a specific method. It refers to the ability of a model to continuously learn and adapt to new data over time, retaining knowledge from previous tasks while learning new ones without forgetting. A model is anticipated to learn from noisy, unpredictable, and shifting data distributions while perpetually accumulating knowledge from previously seen data. The model must transfer previously acquired information to new tasks, transfer new knowledge backward to previously learned tasks, and respond swiftly to contextual changes. Artificial Intelligence

* Corresponding author.
E-mail addresses: as1516645@qu.edu.qa (A.N. Sayed), yhimeur@ud.ac.ae (Y. Himeur), varlamis@hua.gr (I. Varlamis), f.bensaali@qu.edu.qa (F. Bensaali).

https://doi.org/10.1016/j.apenergy.2025.125458
Received 8 October 2024; Received in revised form 27 December 2024; Accepted 25 January 2025
Available online 10 February 2025

Table 1
Comparison of CL surveys and reviews.

Survey	Year	Application domain/s	Research gaps
Parisi et al. [17]	2019	General	• Primarily theoretical, it lacks application-specific insights such as in energy systems.
Hadsell et al. [18]	2020	General	• Broader conceptual discussion; lacks specific domain applications like power systems or energy management.
De Lange et al. [19]	2021	General	• Focuses heavily on classification tasks, with limited exploration of energy-related applications.
Ven et al. [20]	2022	General	• Conceptual framework but lacks empirical validation, particularly in energy systems.
Wickramasinghe et al. [21]	2023	General	General discussion without empirical evaluation or focus on energy systems.
Wang et al. [22]	2024	General	• General review; lacks a focused analysis on real-world energy applications or systems.
Lesort et al. [23]	2020	Robotics, autonomous systems	• Robotics-centric; does not generalize to energy systems or non-robotic domains.
Shaheen et al. [24]	2022	Autonomous systems, robotics	• Emphasis on autonomous systems; lacks detailed insights for energy management.
Mia et al. [25]	2022	Vision, image classification	• Limited to vision tasks; does not address energy system applications or practical deployment.
Hurtado et al. [26]	2023	Predictive maintenance	• Limited to predictive maintenance without addressing broader energy management challenges.
Proposed review	2024	Energy management	• Addresses energy-specific applications of CL, emphasizing power systems, NILM, and smart grids. Highlights unique challenges and opportunities in this domain.

(AI) systems' capabilities have grown significantly recently, but creating a lifetime learning system is still challenging. Addressing this issue requires algorithmic advances.

The main obstacle in CL is retaining information without encountering catastrophic forgetting, also known as catastrophic interference. This means that performance on a previously learned task or domain should not decline considerably over time as additional tasks or domains are introduced. This phenomenon occurs since traditional neural networks tend to override weights associated with previous tasks while training for a new task. This obstacle also stems directly from a more significant problem in neural networks: the stability–plasticity conundrum [22]. Plasticity relates to the capacity to integrate new knowledge, whereas stability refers to retaining old knowledge while dealing with incoming data. Despite its complexities, advances in CL are resulting in the emergence of real-world applications [23,26–28].

CL provides a transformative potential for energy management systems by addressing the need for adaptability in dynamic and non-stationary settings. Modern energy systems face challenges such as fluctuations in demand, seasonal changes, evolving user behaviors, and the introduction of renewable resources. These challenges require the system to learn and adapt in real time. Unlike traditional ML approaches that necessitate retraining on static datasets, CL enables systems to incrementally learn from new data while retaining knowledge from previous tasks, avoiding catastrophic forgetting. This capability is essential for energy systems to maintain optimal performance, adapt to changes, and support long-term operational goals.

1.1. Related surveys

Many surveys have investigated the field of CL. These reviews can be broadly categorized into two main groups based on their thematic content. The first group primarily focuses on CL methods and their general applications, while the second group concentrates on specific applications, discussing CL in the context of those particular use cases. The following papers belong to the first category: Parisi et al. [17] categorized CL approaches into three main types: replay, regularization, and architectural methods. Hadsell et al. [18] discussed the conceptual aspects of CL but did not provide a thorough technical survey. In their survey, De Lange et al. [19] classified CL approaches into replay, regularization, and parameter isolation methods and conducted a case study for further illustration. Ven et al. [20] developed a conceptual framework for CL and practical examples to support their framework. Wickramasinghe et al. [21] identified various CL techniques and the related challenges, offering valuable insights for future research. Lastly, Wang et al. [22] presented an overview of the core principles, techniques, and applications associated with CL.

As for the second category, the following papers are included: Lesort et al. [23] presented a framework that integrates task-based and lifelong learning approaches for robotics applications. Shaheen et al. [24] addressed real-world challenges and deployment considerations for autonomous applications. Mia et al. [25] assessed state-of-the-art methods for vision and image classification tasks across various benchmarks. Hurtado et al. [26] discussed domain-specific challenges and strategies for deploying predictive maintenance in CL applications. Table 1 summarizes the recent reviews in CL to highlight the contributions of this survey. As shown in the table, recent surveys have yet to cover the application of CL in the energy management domain, which is the focus of this survey paper.

1.2. Contribution of the paper

The continuous generation of data by sensors, the introduction of new tasks associated with using new devices or households on a power network, and the evolving consumption profiles over time all contribute to the complexity of energy-related data-driven tasks. These factors highlight the need for CL models over static ones, as CL allows for adapting to changing conditions and incorporating new information into the energy management process. Thus, this research is driven by the scarcity of existing literature focusing on the detailed analysis of energy management systems integrating CL. The primary contributions of this paper can be summarized as follows:

- Providing, to the best of the authors' knowledge, the first review that investigates and summarizes the importance and implementation of CL for energy management systems.
- Conducting an extensive case study comparing various CL methods for anomaly detection in the context of the energy-efficient domain.
- Identifying key challenges that could hinder the deployment of CL in the energy domain and specifying prospective approaches to address those challenges.

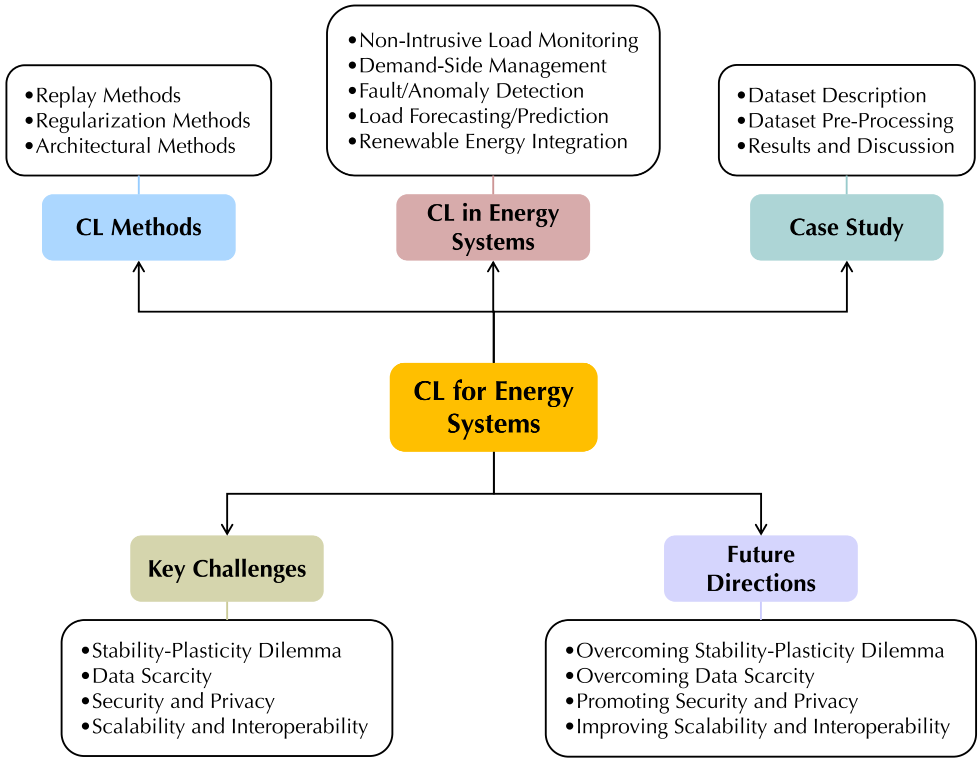

Fig. 1. Taxonomy of CL for power/energy systems.

1.3. Organization of the paper

This review focuses on practical energy-related applications of CL as identified in the literature, using the review methodology found in Section 2. Section 3 provides an overview of the CL techniques and their classifications. Section 4 discusses the practical applications of CL in the energy domains. Section 5 presents a case study that compares CL techniques for anomaly detection. Section 6 discusses key challenges in deploying CL in the energy domain, and future directions are provided in Section 7. Lastly, in Section 8, concluding remarks are made, and potential future directions are outlined.

2. Review methodology

2.1. Study selection

Fig. 1 demonstrates the taxonomy applied in this review to categorize existing studies based on different aspects, including recent CL models, applications of CL in the energy domain, current challenges, and future directions. The review followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) [29] standard, a practical and efficient technique for conducting survey studies.

2.2. Inclusion/exclusion criteria

All selected frameworks have been rigorously reviewed and carefully analyzed based on the inclusion/exclusion method described in the following: (i) frameworks for CL models have been discussed and Table 2 illustrates the search queries used in Scopus database while conducting this review, (ii) only studies published between January 2019 and July 2024 were investigated, (iii) only research publications available online (i.e., peer-reviewed conference proceedings papers, book chapters, and journal articles) were included, (iv) when the same authors published different frameworks on the same problem, the most recent and valuable ones were analyzed, and (v) only studies written in English were considered.

2.3. Quantitative analysis

After conducting the search, numerous studies were found that utilized various CL approaches within the different energy management systems. A quantitative analysis corresponding to the specifics of referenced studies is presented in this subsection. We provide the research statistics on the articles found in Fig. 2 concerning the type of CL approach employed, the CL scenario, where the CL was applied regarding the energy domain, the year of publication, and the first author's affiliated country. In the field of energy, a significant portion of the academic papers that have employed CL approaches have centered their research on two main areas: load forecasting, fault/anomaly detection, and the integration of renewable energy sources. Most CL methods employed in the examined articles were regularization-based, followed by architectural and replay-based techniques. Additionally, the figure depicts the CL scenarios employed in the reviewed papers, with online-based studies appearing most frequently. The figure also illustrates the distribution of publications from 2019 to 2024, highlighting a steady increase in the percentage of papers published each year, with the highest concentration observed in 2023. Lastly, statistics referencing the first author's affiliation country are also provided in the figure. China has the most publications, with 25 papers, followed by South Korea, with eight papers.

3. Overview of CL methods

DL has demonstrated tremendous success in various computer vision and audio processing applications in recent years. However, the primary focus of DL has been on developing high-accuracy Deep Neural Networks (DNN) via offline training with a pre-defined/collected training dataset. The weights in these DNNs are meant to stay static after deployment and do not adapt to changing contexts. Real-world applications, particularly those involving autonomous agents, deal with non-stationary data (i.e., data/tasks that change over time). Therefore, static models fail to perform well in such cases. One option for adapting to changing conditions is to repeat the training process each time

Table 2
Search queries used when conducting the review.

Parameter	Search query
Continual learning	"Lifelong Learning" OR "Sequential Learning" OR "Incremental Learning" OR "Continuously Learning" OR "Continuous Learning"
Energy management system	"Energy System" OR "Energy Systems" OR "Power System" OR "Power Systems" OR "Load Forecasting" OR "Demand Response" OR "Renewable Energy Sources" OR "Presence Detection" OR "Non-Intrusive Load Monitoring" OR "Buildings Energy Consumption" OR "Process Control Applications" OR "Renewable Energies" OR "Energy Management" OR "Energy Storage" OR "Smart Grid" OR "Smart Grids" OR "Microgrid" OR "Microgrid"

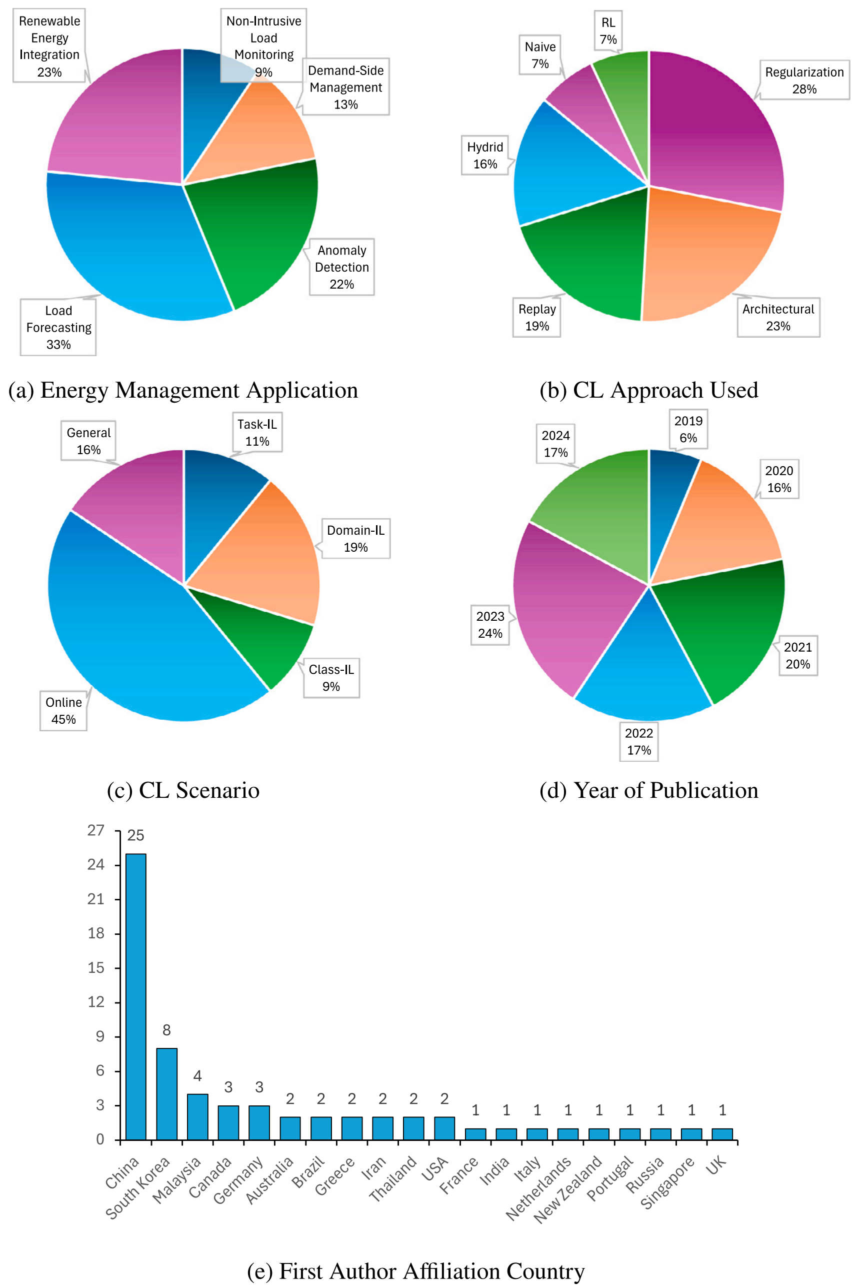

Fig. 2. Statistical findings from the surveyed papers.

a distribution change occurs. However, repeating the entire training procedure, or even for a few epochs, with an enlarged dataset is extremely computationally costly, which makes it impossible in practical resource-constrained circumstances. As a result, it is necessary to develop entirely distinctive techniques/algorithms capable of facilitating resource-efficient continuous learning in real-world systems [24]. CL refers to the ability of a model to learn continuously from a stream of data, adapting to new tasks while retaining knowledge from previously learned tasks. This process involves several mathematical concepts and techniques to ensure the stability and plasticity of the learning model.

Fig. 3 depicts standard update methods for data-driven models with continuous data streams. The continuous data stream can be

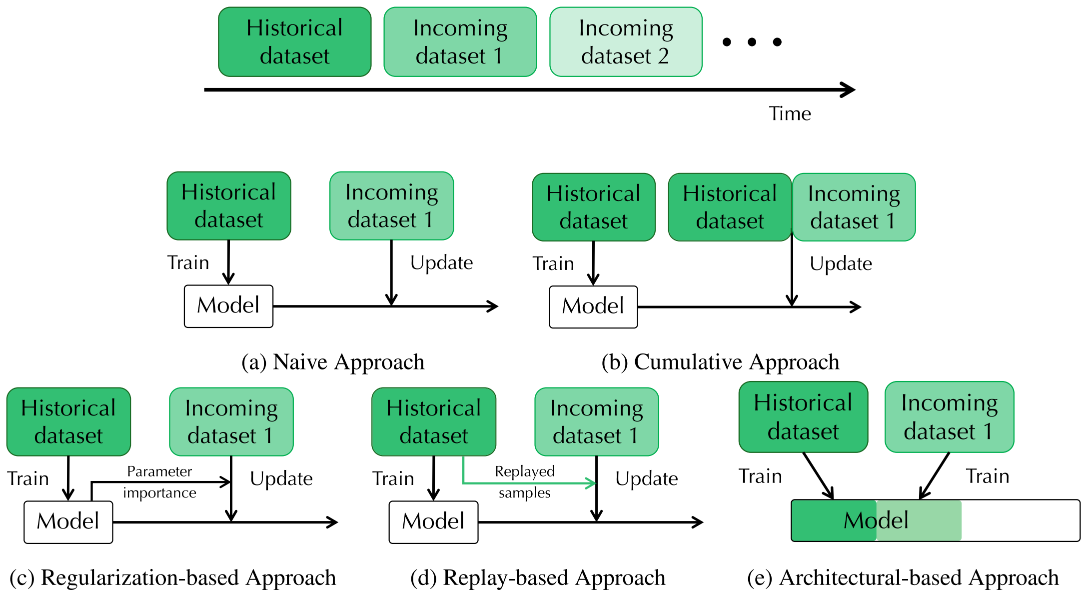

Fig. 3. Standard update approaches with continuous data stream [30].

separated into many subsets (i.e., historical and incoming datasets) based on the timesteps at which the model is updated. The traditional gradient descent-based updating methods (i.e., cumulative learning and naive/fine-tuning learning) face challenges in addressing the concept drift issue. Cumulative learning is time-consuming and requires significant data storage, while fine-tuning cannot prevent catastrophic forgetting. Moreover, Fig. 3(c–e) demonstrates how various CL methods execute model updates.

3.1. CL scenarios

In traditional ML, the model has access to all training data simultaneously. On the contrary, when it comes to CL, the data arrives in batches, and the data distribution shifts over time. Regarding the shifts in data distribution, there are five fundamental ways a supervised learning problem can be incremental [20], as shown in Table 3. The first scenario of CL is known as ‘Task-Incremental Learning’ (or Task-IL). This scenario can be represented as a situation in which an algorithm must learn a series of distinct tasks. A key feature of Task-IL is that the algorithm always knows which task it needs to perform during testing. The second scenario is termed ‘Domain-Incremental Learning’ (or Domain-IL). In this scenario, the overall structure of the problem remains constant, but there are variations in the context or input distribution (for instance, domain shifts). Similar to task-IL, this scenario can also be viewed as an algorithm that must learn a set of ‘tasks’ incrementally (though it may be more accurate to consider them as ‘domains’), with the important distinction that—at least during testing—the algorithm does not know which task a sample corresponds to. Nonetheless, recognizing the task is not required since each task has the same potential outputs (e.g., the same classes are applicable in each task). The third scenario of CL is referred to as ‘Class-Incremental Learning’ (or Class-IL). This scenario is best expressed as one in which an algorithm needs to learn to differentiate among an expanding number of objects or classes over time. A commonly used setup for this situation involves encountering a sequence of classification-based tasks, where each task presents different classes, and the algorithm has to learn to differentiate among all the classes.

The final two types of CL scenarios are online and general CL. Online CL (also called Instance-Incremental Learning (Instance-IL)) focuses on processing data as it arrives, typically one instance or batch at a time, without the ability to revisit past data. This setup mimics real-time learning environments, such as edge devices or Internet of Things (IoT) systems, where data is continuously streamed and memory or computational resources are highly constrained. Finally, general CL aims to handle diverse and complex real-world scenarios where task boundaries, data distributions, and learning objectives are unclear. This approach seeks to unify task, domain, and class-IL elements while accommodating changing data characteristics, overlapping tasks, and real-world unpredictability. The primary distinction between these two approaches is that it is ambiguous which CL scenarios were applied in studies utilizing online learning. On the other hand, the studies using two or more CL scenarios simultaneously are categorized as having a general learning scenario.

3.2. Catastrophic forgetting

Catastrophic forgetting is one of the major problems in CL. Catastrophic forgetting occurs when a model trained on new data significantly degrades its performance on previously learned tasks. Mathematically, if $\mathcal{T}_i$ represents the ith task, the model parameters θ are optimized for the current task $\mathcal{T}_n$, and the loss function for the current task is $\mathcal{L}_{\mathcal{T}_n}(\theta)$, the training objective can be written as [32]:

$$\theta^* = \arg\min_{\theta} \mathcal{L}_{\mathcal{T}_n}(\theta) \tag{1}$$

In CL, the goal is to optimize θ such that the loss on previous tasks is also minimized:

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{n} \mathcal{L}_{\mathcal{T}_i}(\theta) \tag{2}$$

In principle, recent research attempts to tackle catastrophic forgetting during continuous learning, using longer task sequences with more samples. Based on how task-specific information is stored and used throughout the sequential learning process, the following approaches are distinguished in CL literature [19]:

1. Regularization Methods,
2. Replay Methods, and
3. Architectural Methods

3.3. Regularization methods

This method prioritizes privacy and reduces memory needs by avoiding using raw inputs. Rather, while learning with new data, an additional regularization term is added to the loss function to consolidate

Table 3
Overview of CL scenarios.

Scenario	Description
Task-Incremental (Task-IL) [20]	• A new task with its own set of classes is introduced incrementally. Tasks are typically distinguished by a task ID.
Domain-Incremental (Domain-IL) [20]	• The same set of tasks or classes are learned under changing conditions, such as shifts in data distribution or context.
Class-Incremental (Class-IL) [20]	• New classes are introduced incrementally, and the model must learn them without forgetting previously learned classes.
Online Learning/Instance-Incremental Learning (Instance-IL) [22]	• Data is processed one instance or batch at a time without revisiting previous data.
General/Hybrid CL [31]	• Combines multiple scenarios or focuses on real-world setups where task boundaries and data distribution changes are unknown.

prior information [33–36]. These techniques may be further separated into two categories: data-focused and prior-focused. Regularization methods constrain the optimization problem to preserve knowledge from previous tasks.

3.3.1. Elastic weight consolidation (EWC)

EWC adds a regularization term that penalizes changes to important parameters. The importance of each parameter θ_j is measured by the Fisher Information Matrix F [37]:

$$\mathcal{L}_{\text{EWC}}(\theta) = \mathcal{L}_{\mathcal{T}_n}(\theta) + \frac{\lambda}{2}\sum_j F_j(\theta_j - \theta_j^*)^2, \tag{3}$$

where θ_j^* are the optimal parameters from previous tasks and λ is a regularization strength hyperparameter.

3.3.2. L2 regularization

L2 regularization penalizes large changes in the parameter values [36]:

$$\mathcal{L}_{\text{L2}}(\theta) = \mathcal{L}_{\mathcal{T}_n}(\theta) + \frac{\lambda}{2}\sum_j (\theta_j - \theta_j^*)^2 \tag{4}$$

3.3.3. Knowledge distillation (KD)

KD involves training the model to mimic the outputs of the previous model. A well-known CL approach utilizes this concept, which is Learning without Forgetting (LwF) [38]. The distillation loss $\mathcal{L}_{\text{KD}}$ is defined as [39]:

$$\mathcal{L}_{\text{KD}}(\theta) = \sum_{x\in D_{\text{old}}} D_{\text{KL}}(p_{\text{old}}(y|x)\|p_{\text{new}}(y|x,\theta)) \tag{5}$$

where D_{KL} is the Kullback–Leibler divergence, p_{old} is the output distribution of the old model, and p_{new} is the output distribution of the new model.

3.4. Replay methods

In the replay methods, also commonly referred to as rehearsal methods, raw samples from previously seen tasks or pseudo-samples produced using a generative model are stored in the system memory. These prior task samples are repeated during model training to reduce forgetting while learning a new task. This approach either reuses the stored instances as model inputs for rehearsing or constrains the optimization of the current task loss to avoid interference from earlier tasks. This method could be sub-categorized into three branches, namely, (1) rehearsal, (2) pseudo rehearsal, and (3) constrained methods.

3.4.1. Experience replay

Experience replay stores a buffer $\mathcal{B}$ of past data samples and mixes them with current data during training [40]. Due to the relatively restricted storage capacity, the primary problem is designing and leveraging the memory buffer. In terms of construction, the stored training samples should be carefully picked, compressed, supplemented, and updated to adaptively retrieve previous knowledge:

$$\mathcal{L}_{\text{ER}}(\theta) = \mathcal{L}_{\mathcal{T}_n}(\theta) + \sum_{(x,y)\in\mathcal{B}} \mathcal{L}_{\text{old}}(x,y;\theta) \tag{6}$$

3.4.2. Generative replay

Pseudo-rehearsal (also known as generative replay) methods train generative AI models to synthesize pseudo-examples that fit previous tasks, which are then used for rehearsal [41]. This approach is particularly useful in scenarios where samples from previous tasks are unavailable or not stored by the system. This is accomplished by leveraging the outcomes of the prior tasks to train the generative model instead of storing them in memory. The generated random samples are then periodically fed into the latest task-driven trained model to avoid catastrophic forgetting. As new tasks appear, more samples are used to fine-tune the generative model and provide comparable output results (pseudo-samples) [42]. Generative replay uses a generative model G to produce synthetic data for previous tasks [43]:

$$\mathcal{L}_{\text{GR}}(\theta) = \mathcal{L}_{\mathcal{T}_n}(\theta) + \sum_{(x,y)\in G} \mathcal{L}_{\text{old}}(x,y;\theta) \tag{7}$$

3.4.3. Constrained replay

Experience replay methods are widely used in ML, particularly Reinforcement Learning (RL) and CL. However, these approaches often need help with significant challenges, primarily due to their tendency to overfit the examples stored in their replay buffers. This overfitting can result in degraded performance compared to models trained in a more integrated manner, where all data is available concurrently. To address this issue, researchers have explored various constrained replay strategies to mitigate the negative impact of overfitting while maintaining the learning from previous tasks.

One notable approach in this realm is Gradient Episodic Memory (GEM) [44]. GEM allows for integrating updates from new tasks without disrupting the knowledge acquired from previous tasks. It accomplishes this by projecting the gradients of new task updates onto a feasible region defined by the gradients from earlier tasks. This projection is facilitated through the use of a first-order Taylor series approximation, ensuring that the learning process is both efficient and preserves prior knowledge effectively. Another significant variant within constrained replay methods is Averaged GEM (AGEM) [45]. AGEM simplifies the problem by allowing projections into a single direction determined by a randomly sampled subset from the buffer of previous task data. This method relaxes the constraints, enabling the model to retain the valuable information from older tasks while remaining flexible enough to incorporate new data. By leveraging this strategy, AGEM effectively balances the retention of past knowledge with acquiring new skills.

3.5. Architectural methods

To help with forgetting, this family of approaches assigns distinct model parameters to each task. This is done by creating new branches for new tasks while freezing the parameters of earlier activities or assigning a model copy to each task when there are no restrictions on the size of the architecture. Architectural methods dynamically adjust the model's architecture to accommodate new tasks.

3.5.1. Progressive neural networks (PNNs)

PNNs add new subnetworks for each task and keep previously trained subnetworks fixed [46]:

$$\theta = \{\theta_1, \theta_2, \dots, \theta_i\}, \tag{8}$$

where θ_i are the parameters of the ith subnetwork.

3.5.2. PackNAT

This method [47] allocates parameter groups to subsequent tasks iteratively using binary masks. New tasks are designed to establish two training stages. First, the network is trained without modifying the prior task parameter groups. Then, specific insignificant free parameters are trimmed and monitored using the lowest magnitude. The remaining subset of essential parameters is retrained during the second training phase. The pruning mask maintains task performance by guaranteeing that the task parameter subset remains constant for subsequent tasks. PackNet provides explicit network bandwidth allocation per job, restricting the number of functions.

3.5.3. Hard attention to the task (HAT)

HAT [48] undergoes a single training phase and includes task-specific embeddings to mask attention. A Sigmoid function controls each layer's embeddings to create attention masks during the forward pass. The Sigmoid slope is adjusted during each training epoch, initially allowing modifications to the masks and eventually resulting in nearly binary masks. A regularization term enforces sparsity on the new task attention mask to accommodate additional tasks. The critical aspect of this approach is restricting parameter updates between two units that are important for previous tasks based on the attention masks.

3.6. Analysis of the CL methods

In summary, *regularization-based methods*, such as EWC, combat catastrophic forgetting by incorporating penalty terms into the loss function. This approach prevents significant updates to crucial parameters for previously learned tasks. These computationally efficient and memory-friendly methods make them suitable for resource-constrained environments. However, they become less effective with highly diverse or complex tasks, as accurately estimating the importance of parameters over time can be challenging. Although they help maintain stability, regularization techniques often only partially mitigate task interference, and their effectiveness may diminish as the number of tasks increases.

Replay methods, while particularly effective in scenarios with ample storage and computational power, also present challenges, such as potential data privacy concerns and the need for strategic sampling to balance memory constraints while maintaining task diversity. Generative replay addresses some storage issues but heavily depends on the quality of generated samples, which can decline over time, especially in high-dimensional data spaces.

On the other hand, *architecture-based methods* enhance a model's structure by adding new units, layers, or subnetworks as new tasks are introduced. This strategy prevents task interference and ensures that previously learned knowledge is preserved. These methods offer task isolation and long-term scalability, making them ideal for systems managing multiple tasks without performance degradation. However, the rapid growth of the model size can lead to inefficiencies in computation and memory usage. Additionally, managing and training a growing architecture can become increasingly complex, making these methods more computationally expensive and challenging to deploy in resource-constrained settings.

To enhance learning efficiency, hybrid approaches merge elements of regularization, replay, and architectural strategies, leveraging the strengths of multiple techniques. For instance, one approach would be to pair generative replay with regularization to balance stability and flexibility. While these methods show promise, they often encounter increased computational complexity, making them less practical in resource-constrained environments.

3.7. Evaluation metrics

Several metrics are used to evaluate CL models:

3.7.1. Accuracy (A)

The accuracy of a specific task or set of tasks is defined as:

$$A = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{9}$$

3.7.2. Average accuracy (AA)

The average accuracy at task n is then defined as:

$$AA_n = \frac{1}{n}\sum_{i=1}^{n} a_{n,i}, \tag{10}$$

where $a_{n,i} \in [0,1]$ is the accuracy assessed on the test set of the ith task $(i \le n)$ after training the network incrementally from tasks 1 to n.

3.7.3. Backward transfer (BWT)

The influence of learning new tasks on the performance of previously learned tasks is defined as [49]:

$$BWT = \frac{1}{n-1}\sum_{i=1}^{n-1}(A_i - A_i^*), \tag{11}$$

where A_i^* is the accuracy on task i immediately after learning it, and A_i is the accuracy on task i after learning the new tasks.

3.7.4. Forgetting (F)

The reduction in performance on previous tasks after learning new tasks is defined as [50]:

$$F = -BWT \tag{12}$$

3.8. Synopsis of recent CL research

CL has emerged as a pivotal area in ML, aiming to enable models to learn from non-stationary data streams while mitigating catastrophic forgetting and preserving knowledge from previous tasks. Recent research spans diverse methodologies, leveraging benchmarks, pre-trained models, task-specific optimizations, generative techniques, and biologically inspired approaches. This subsection overviews notable advancements and innovative solutions across these dimensions, highlighting key contributions and their impact on the field. It is worth noting that most existing research on CL has focused on specific methodologies and performance evaluations using well-known datasets like MNIST or CIFAR-100 [20].

3.8.1. CL benchmarks and evaluation

Several works highlight innovative approaches to tackling key challenges in CL across diverse applications and domains. For example, Verwimp et al. [51] introduced the CLAD benchmark for autonomous driving, utilizing the SODA10M dataset to address classification and detection tasks with domain-IL challenges while highlighting the limitations of existing CL benchmarks and suggesting pathways for future research. Ghunaim et al. [52] emphasized the importance of computational efficiency in CL with their real-time evaluation framework, demonstrating that a simple baseline outperformed state-of-the-art methods on the large-scale CLOC dataset. Similarly focused on improving CL performance, Razdaibiedina et al. [53] proposed Progressive Prompts for language models, achieving over 20% improvements in test accuracy on the T5 model compared to prior state-of-the-art methods. Complementing these efforts, Smith et al. [54] explored rehearsal-free CL, showing that L2 parameter regularization combined with K provided superior performance on benchmarks such as ImageNet-R and CIFAR-100, outperforming other techniques while addressing catastrophic forgetting.

3.8.2. Approaches leveraging pre-trained models

Various studies highlight innovative strategies for harnessing pre-trained models and prompting techniques to advance CL across diverse domains. For instance, Mcdonnell et al. [55] leveraged pre-trained models for CL by introducing a random projector and class-prototype accumulation approach, effectively reducing error rates by 20%–62% across benchmarks without relying on rehearsal memory. Expanding on the use of pre-trained models, Gao et al. [56] proposed the Learning-Accumulation-Ensemble (LAE) framework, integrating Parameter-Efficient-Tuning (PET) methods to achieve accuracy improvements of 1.3% on CIFAR-100 and 3.6% on ImageNet-R. Focusing on video data, Villa et al. [57] introduced PIVOT, a prompting-based method that utilized pre-trained image models to achieve a significant 27% improvement on the ActivityNet benchmark. Complementing these efforts, Smith et al. [58] developed an end-to-end attention-based key-query prompting mechanism, which surpassed the performance of DualPrompt with up to a 4.5% accuracy increase on established CL benchmarks.

3.8.3. Task-specific CL

Various frameworks demonstrate advancements in CL methodologies by tackling core challenges such as pre-training optimization, class imbalance, and the stability–plasticity trade-off. Typically, Wang et al. [59] introduced HiDe-Prompt, a hierarchical approach for optimizing pre-training in CL, achieving significant improvements of 15.01% on Split CIFAR-100 and 9.61% on Split ImageNet-R. Building on strategies to enhance CL performance, Lin et al. [60] proposed Proxy-based Contrastive Replay (PCR), effectively addressing class imbalance and instability in replay-based methods, yielding superior results on benchmarks such as Split CIFAR-100 and Split MiniImageNet. Complementing these efforts, Kim et al. [61] presented Auxiliary Network CL (ANCL), which improved the trade-off between plasticity and stability, achieving 1%–3% better accuracy in IL scenarios.

3.8.4. Generative and diffusion-based methods

Numerous studies propose innovative approaches to improve functionality and address challenges in generative and vision-language models within the realm of CL. Specifically, Heng et al. [62] introduced Selective Amnesia, a novel approach for controllable forgetting in text-to-image models, enabling the removal of sensitive content while preserving overall functionality. Expanding on generative methods, Gao et al. [63] proposed Deep Diffusion-based Generative Replay (DDGR), which significantly enhanced generative replay approaches, achieving a 4.71% accuracy improvement on CIFAR-100. Addressing zero-shot transfer challenges in vision-language models, Zheng et al. [64] developed Zero-Shot CL (ZSCL), a method to prevent performance degradation, leading to a notable 9.7% improvement on a multi-domain task benchmark.

3.8.5. Biological and regularization-inspired approaches

To address fundamental challenges in CL, some studies have adopted innovative approaches inspired by biological systems and advanced algorithms. In this regard, Wang et al. [65] drew inspiration from the Drosophila learning system to develop a model that effectively balances plasticity and stability, outperforming traditional synaptic regularization methods in task-IL settings. Building on the challenge of maintaining plasticity, Dohare et al. [66] introduced continual backpropagation, a novel approach designed to counter the loss of plasticity in CL scenarios, demonstrating notable improvements on MNIST and ImageNet tasks.

In Table 4, we aim to provide a summary of the above-described studies based on different aspects, including specific models employed, the datasets used or types of data, the real-world applications of the research, the primary contributions of each study, the performance metrics used, and any limitations or drawbacks identified in the research.

3.9. Implementing CL

The implementation of CL begins with defining the problem and tasks, followed by choosing a suitable CL method from options such as regularization-based methods, replay methods, parameter isolation methods, or dynamic architectures. The model is then initialized to prepare it for the training phase. Moving on, for each task in the sequence, the algorithm pre-processes the data, trains the model on the current task, and applies techniques to address catastrophic forgetting, ensuring previously learned knowledge is retained. Methods for mitigating forgetting include KD, EWC, generative replay, and using a memory buffer. After all tasks have been processed, the model is evaluated to assess its performance, and further optimization and refinement steps are applied to enhance the model's efficiency and accuracy. This comprehensive approach ensures the model continuously learns and adapts while preserving knowledge from previous tasks. Algorithm 1 outlines a structured approach to implementing CL.

Algorithm 1: Steps to Implement CL

Data: Tasks $T_1, T_2, \ldots, T_n$
Result: Trained CL model
Function `Main`($tasks$):
    Define the Problem and Tasks;
    strategy ← ChooseContinualLearningStrategy();
    model ← ModelInitialization();
    foreach *task* $T_i \in tasks$ **do**
        PreprocessData(T_i);
        TrainModel($model, T_i$);
        HandleCatastrophicForgetting($model$);
    EvaluateModel($model$);
    OptimizeAndRefine($model$);
Function `ChooseContinualLearningStrategy()`:
    return Regularization-Based Methods, Replay Methods, Parameter Isolation Methods, Dynamic Architectures;
Function `HandleCatastrophicForgetting`($model$):
    return Knowledge Distillation, Elastic Weight Consolidation, Generative Replay, Memory Buffer;
Function `TrainModel`($model, task$):
    PreprocessData($task$);
    TrainModelAlgorithm($model, task$);
    ValidateAndTune($model, task$);
Function `OptimizeAndRefine`($model$):
`Main`($T_1, T_2, \ldots, T_n$)

Table 4
Comparison of studies on CL.

Ref.	Model(s) used	Application	Dataset/Data type	Main contribution	Best performance value	Limitation
[51]	Various CL models	Object classification and detection in autonomous driving	SODA10M	Introduced CLAD-C (classification) and CLAD-D (detection) benchmarks	mAP: 59.8 (Finetune), 74.7 (D1), 61.5 (D2), 59.0 (D3)	Focuses on specific benchmarks for autonomous driving
[52]	Various CL models	Real-time evaluation of CL	CLOC	Evaluated CL methods with computational cost constraints	ER outperformed other methods under varying memory budgets	Existing methods fail in practical settings due to computational costs
[55]	Pre-trained models with random projectors	CL with pre-trained models	Various class-IL benchmarks	Proposed a training-free approach using random projectors and class-prototype accumulation	Reduced final error rates by 20%–62% on seven class-IL datasets	Focuses on class- and domain-IL CL without rehearsal memory
[67]	Online prototype learning (OnPro)	Online CL	CIFAR-10, CIFAR-100	Introduced OnPro framework to tackle shortcut learning in online CL	Acc: 57.8% (CIFAR-10), 22.7% (CIFAR-100)	Limited to online CL scenarios
[68]	Prompt-based methods with language guidance	CL without replay buffer	Various datasets for prompt-based methods	Proposed LGCL for improving prompt-based methods	Avg. Accuracy: 86.15 (Frozen CLIP Keys), 87.23 (Learnable Keys)	Focuses on prompt-based methods with language guidance
[65]	Multiple learning modules inspired by Drosophila	General CL	Various datasets	Proposed a multi-learner architecture inspired by biological systems	Improved performance over synaptic regularization methods	Complex architecture may be difficult to implement
[69]	Pre-trained models with Slow Learner	CL on pre-trained models	Split CIFAR-100, Split ImageNet-R, Split CUB-200, Split Cars-196	Proposed SLCA to address progressive overfitting in CLPM	Up to 49.76% (Split CIFAR-100), 50.05% (Split ImageNet-R)	Focuses on addressing overfitting in CL with pre-trained models
[56]	PET methods	Unified CL framework	CIFAR-100, ImageNet-R	Proposed LAE framework for CL with PET methods	Last-incremental accuracy improved by 1.3% (CIFAR-100), 3.6% (ImageNet-R)	Focuses on parameter-efficient tuning methods
[57]	Prompting mechanisms	CL for video data	ActivityNet	Introduced PIVOT for video data CL using image domain pre-trained models	Improved state-of-the-art by 27% on 20-task ActivityNet setup	Limited to video data CL
[58]	Vision transformer models with key-query mechanism	Computer vision	Various class-IL and domain-IL benchmarks	Proposes an end-to-end attention-based key-query scheme with input-conditioned prompts	4.5% improvement in average final accuracy over DualPrompt	Reduced plasticity, sacrificing new task accuracy
[59]	HiDe-Prompt with task-specific prompts and contrastive regularization	CL with pre-trained knowledge	Split CIFAR-100, Split ImageNet-R	Introduces Hierarchical Decomposition (HiDe-)Prompt	15.01% improvement on Split CIFAR-100, 9.61% on Split ImageNet-R	Difficulty incorporating task-specific knowledge
[62]	Selective Amnesia applied to conditional variational likelihood models	Selective forgetting in deep generative models	Various models and datasets	Enables controllable forgetting of specific concepts	Effective forgetting of harmful concepts in text-to-image models	Risk of forgetting important information
[60]	PCR	Online class-IL CL	Split CIFAR-10, Split CIFAR-100, Split MiniImageNet	Combines proxy-based and contrastive-based replay methods	Final accuracy on Split CIFAR-10: 58.8%, Split CIFAR-100: 29.3%, Split MiniImageNet: 28.4%	Unstable and hard to converge with limited samples

(*continued on next page*)

Table 4 (*continued*).

[64]	ZSCL with parameter and feature space regularization	Vision-language models	Multi-domain Task-IL (MTIL)	Prevents zero-shot transfer degradation	9.7% improvement in MTIL benchmark	Trade-off between zero-shot performance and downstream task performance
[61]	ANCL	Task-IL and class-IL	Various datasets	Promotes plasticity with auxiliary network	1%–3% improvement over naive CL approaches	Underlying mechanism of plasticity-stability trade-off not fully understood
[66]	Continual backpropagation with L2-regularization	CL	MNIST, ImageNet	Introduces continual backpropagation to maintain plasticity	Maintains high accuracy over long task sequences	Performance drops in highly challenging settings
[63]	DDGR	Class-IL	CIFAR-100	Uses diffusion model for generative replay	Final average accuracy of 59.20% on CIFAR-100	Challenges with a large first task and diffusion steps
[53]	Progressive Prompts for language models	CL in language models	Standard CL benchmarks	Sequentially concatenates new soft prompts	>20% improvement over previous best on T5 model	Relies on data replay and task-specific parameters
[54]	L2 parameter regularization with KD	CL without rehearsal	CIFAR-100, ImageNet-R	Combines parameter regularization and KD for strong performance	Final accuracy of 35.6% on CIFAR-100 with pre-training	Large performance gap between rehearsal and rehearsal-free methods
[70]	Drift activated rehearsal	Online class-IL	CIFAR-10, MNIST	Combines drift detection with various rehearsal techniques for optimum performance	Final accuracy 94.9% at MNIST, 80.5% at CIFAR-10	Relies on data replay and task-specific parameters

4. CL in the energy domain

The preceding section provided a comprehensive overview and analysis of the various classes and methodologies used to address CL. This section thoroughly examines the research studies incorporating CL in the energy and power domain. These studies are further classified into five distinct subsections based on the specific areas within the power system where CL was applied. These subsections include:

1. Non-Intrusive Load Monitoring
2. Demand-Side Management
3. Fault/Anomaly Detection
4. Load Forecasting/Prediction
5. Renewable Energy Integration

4.1. Non-intrusive load monitoring

Non-Intrusive Load Monitoring (NILM) uses an algorithmic approach to monitor the states of appliances and their power consumption within a building using a single metering point. Recently, DNN methods have proven to be the most effective for NILM in classifying appliance states as single or multi-label networks. However, most current DNN methods are static and do not consider changes in user habits or appliances. Fig. 4 demonstrates how class-IL can be applied for NILM. The studies referenced have introduced a novel solution integrating CL to address this challenge.

The study in [71] proposed integrating NILM with CL. Specifically, they have compared the proposed Appliance IL (AIL) approach with the LwF approach to adapt and monitor new appliances. Additionally, they compared their technique to a static NILM method, and the findings showed that the suggested strategy effectively copes with newly added appliances while reducing forgetting of previously trained/seen appliances. Similarly, Sykiotis et al. [72] employed a CL approach to perform NILM to address the issue of forgetting previously learned information. To elaborate, the experience replay method was employed to improve NILM model retraining in a resource-efficient manner. Furthermore, NILM was implemented in an IL setting in [73]. The strategy involved using a particular approach to work on the current task with new appliance classes and another method for older appliance classes to identify the associations between the older and newer classes. Moreover, Li et al. [74] proposed a novel IL approach to address catastrophic forgetting in NILM, enabling adaptive learning for new appliance classes and demonstrating effectiveness using the PLAID dataset. Similarly, Zhang et al. [75] presented an IL approach for NILM, enabling progressive identification of new appliances with limited training data. Lastly, Qiu et al. [76] introduced a novel method combining class-IL and semi-supervised learning to address the challenge of accurate load identification in NILM. The method prevents catastrophic forgetting, distills knowledge, and exploits unlabeled data. Experimental results on PLAID and WHITED datasets demonstrated adequate performance.

In NILM, CL addresses key challenges such as device variability, changing user habits, and noisy energy signals. Unlike traditional models, CL enables IL, allowing adaptation to new appliances or usage patterns without full retraining while mitigating catastrophic forgetting through techniques like EWC or experience replay. Through resource efficiency, CL can be ideal for real-time deployment on edge devices, improving accuracy and reducing disaggregation errors. Future directions include tackling imbalanced energy datasets and integrating CL with Federated Learning (FL) for privacy-preserving, adaptive energy monitoring in real-world settings.

4.2. Demand-side management

Demand-side management presents the potential to drastically lower building operating expenses and overall energy usage [77]. Buildings may reduce their energy use and increase their energy efficiency by implementing various management techniques. Enhancing energy efficiency through better materials, intelligent energy tariffs that offer incentives for particular consumption patterns, and sophisticated real-time control of distributed energy resources are

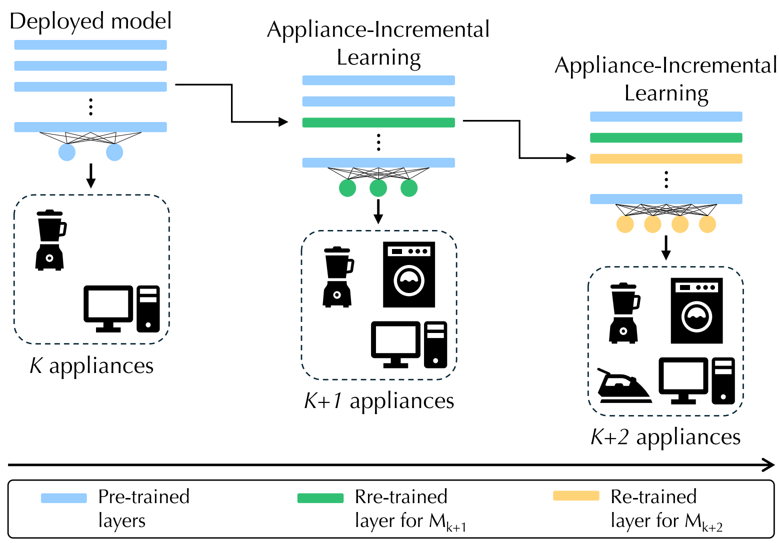

Fig. 4. Overview of the AIL for NILM employing class-IL by introducing a new appliance throughout each re-training round [71]

just a few examples. The studies summarized utilized CL in their implementations.

Kim et al. [78] proposed a demand response approach to controlling and adjusting energy consumption loads such as those from lights and air conditioners in domestic buildings. A data-driven ML approach has tackled the uncertainty in environmental variables and user preferences. Moreover, a Multi-Task Learning (MTL) approach was employed to take advantage of the shared structural characteristics within policies for different rooms. A method for kernel-based lifelong learning has been developed, which can continuously update the shared representation of the policies online to improve computational efficiency and the ability to track optimal policies over time [79]. An overview of the proposed CL-based demand response is illustrated in Fig. 5. Similarly, Hossain et al. [80] employed Random Forest (RF) to determine the demand response schedule. They used the MuZero RL method, which allows for continuous learning through self-play, attains better sample efficiency, and effectively adjusts to changing environments. The MuZero RL was mainly used to manage the demand response, which can reduce energy costs and smooth out peak loads more efficiently. Equivalently, the authors in [81] proposed a Deep RL (DRL) approach to managing energy usage in a cluster of houses. To address the unpredictability of power usage in homes, they used a DNN to analyze household data, including power usage patterns, indoor temperature, outdoor temperature, and humidity. This data was then used to predict the total peak power demand. The DRL approach effectively reduced total power usage and minimized consumption during high-demand periods while maintaining a specific temperature. A demand prediction-based scheduling approach was proposed in [82] using CL and DL. The scheduling algorithm uses the cosine similarity of the electric load pattern to manage and control the gradient of the optimization process. Experimental evaluations demonstrated the effectiveness of the proposed scheme compared to the base method. Similarly, Wu et al. [83] proposed a replacement learning-based online adaptation framework for multivariate multi-step time series forecasting. This approach addressed concept drift through retraining, clustering-based sampling, and a correcting factor, outperforming offline models by over 50% in forecasting accuracy, as validated using synthetic building and cooling system electricity demand datasets.

Tang et al. [84] introduced a demand-aware pricing algorithm using DRL. The algorithm was designed to optimize energy consumption and provide real-time pricing. The algorithm adjusted pricing structures dynamically through continuous learning and adaptation to encourage demand-side flexibility and maintain grid stability. In a similar manner, the authors in [85] proposed a CL-based bidding system for energy trading to address the change in data distribution while using ML. The framework used a CL method, which involved combining a small segment of historical data with new data to enhance the accuracy of bidding decisions. The framework was tested using an energy trading dataset and showed improved prediction accuracy. Lastly, W. Y. Ng et al. [86] introduced an incremental ensemble learning method for streaming data environments, tackling concept drift and class imbalance. The technique employed dynamic cost-sensitive weighting and imbalance-reversed bagging to enhance classification accuracy. Applied to electricity pricing prediction in Australia, the approach demonstrated statistically significant effectiveness compared to state-of-the-art IL techniques, validating its utility in smart grid applications. The Radial Basis Function Neural Network (RBFNN) was employed as the base classifier.

All in all, CL plays a crucial role in demand-side management by enabling energy systems to adapt to evolving consumption patterns and grid dynamics. Unlike traditional models, CL allows for incremental updates, efficiently learning new trends such as shifts in user behavior, the addition of new energy devices, or seasonal demand fluctuations without forgetting previously learned knowledge. CL's lightweight nature also makes it suitable for real-time implementation on edge devices, facilitating smarter energy optimization. Future research could focus on enhancing CL's scalability for large-scale demand-side management applications and improving its integration with predictive models for more accurate demand forecasting.

4.3. Fault/anomaly detection

Fault or anomaly detection in power systems involves identifying unusual occurrences that may indicate defects, malfunctions, or inefficiencies within electrical networks or equipment. This process is vital for ensuring the reliability and safety of power systems, as it allows for the early detection of problems, prevents potential damage, and minimizes downtime. The following papers have incorporated CL into the fault detection process.

Transit Stability Assessment (TSA) ensures the power system's stable and safe operation, particularly after a system disturbance. The authors in [88] merged the advantages of Convolutional Neural Networks (CNNs) and the CL-based algorithm named Orthogonal Weight Modification (OWM) to take advantage of CNN's feature extraction property and OWM's LwF. The proposed approach enabled the evaluation of the system's transient stability and seamless updating per the power grid's configuration. Likewise, in their study, Liu et al. [89] suggested using Support Vector Machines (SVM) with Karush–Kuhn–Tucker (KKT) for

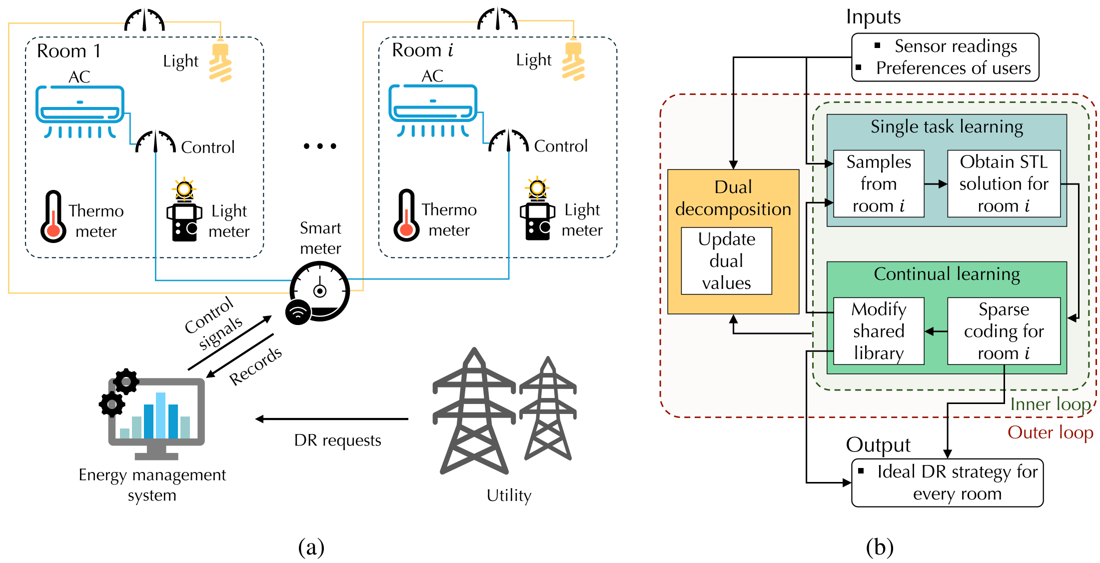

Fig. 5. An instance of applying CL for demand response, where (a) is an overview of the system's architecture and (b) is the proposed CL-based approach to demand response [78].

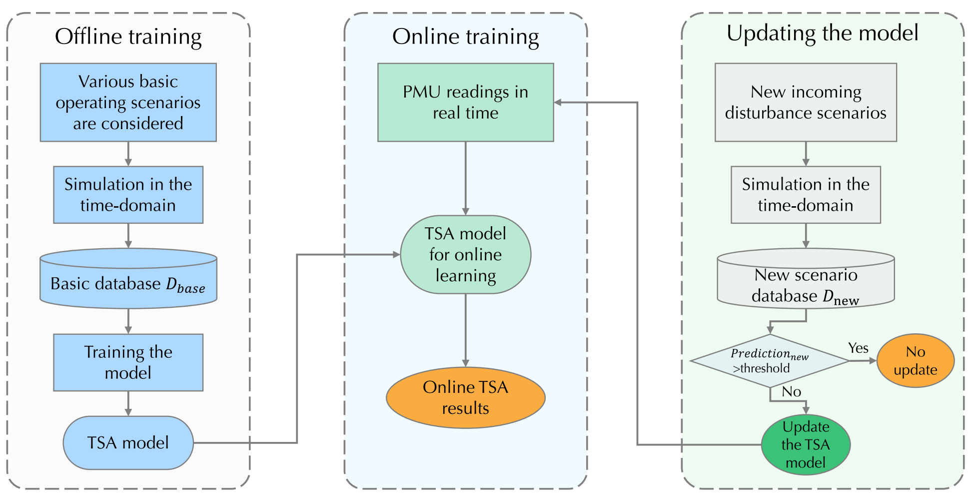

Fig. 6. Overview of CL-based TSA in power systems [87].

performing TSA in an IL setting. To test their approach, they conducted a simulation study using an IEEE-39 bus system. The results indicated that the TSA was updated effectively, and the training was completed efficiently while maintaining prediction accuracy. Moreover, [87] introduced using the Sliced Cramer Preservation (SCP) modal as a CL approach. They combined the SCP algorithm with the Deep Residual Shrinkage Network (DRSN) to create a TSA classifier. Using SCP, the model can be extended and updated using new scenario data. This updated model improves prediction accuracy for new scenarios and maintains prediction capabilities for old scenarios, thus reducing the need for frequent model updates. Furthermore, Ren et al. [90] developed a fully data-driven method for post-fault short-term voltage stability assessment, addressing missing Phasor Measurement Units (PMU) measurements using a deep residual learning CNN and Incremental Broad Learning (IBL). Similarly, Tian et al. [91] proposed a novel transient stability boundary construction methodology for power systems, modeling critical clearing times using a Broad Learning System (BLS). This efficient, update-capable method, validated through IEEE and real-world case studies, demonstrated superior accuracy and robustness in online transient stability assessment. Also, Li et al. [92] introduced a dual cost-sensitivity factor method for TSA to address class and regional imbalances. Using LightGBM, the approach improved unstable sample accuracy and reduced stable sample misjudgment. As validated on three power systems, an IL-based fast update scheme further enhanced online performance. Additionally, Cui et al. [93] proposed a TSA digital twin framework using KD to address computational resource challenges. This approach enabled adaptive updates for complex conditions, enhancing the model's learning capability. The approach, validated on the IEEE 39-bus system, demonstrated its effectiveness in real-time power grid applications. Fig. 6 showcases how CL could be applied to TSA. The figure highlights the innovative approach of integrating CL techniques to systematically analyze and improve the stability of transit systems over time.

On another note, the study in [94] presented an approach to detect abnormalities in a PMU data stream. The proposed approach consists of an offline and online Gaussian Mixture Model (GMM) to identify if anomalies are present and an ensemble clustering model to classify the present anomaly. The online GMM model was used to update the overall model in the case of concept draft while considering the previous knowledge and alleviating catastrophic forgetting. Sifat et al. [95] proposed a data-driven approach to detect high-impedance faults, a paramount stage in power system distribution. The developed hybrid model was constructed using a CNN phase and several Recurrent Neural Network (RNN) algorithm-based versions. The models were trained using real-world Giant Magneto-Resistive (GMR) device data from a purpose-built 400-volt test facility. A memory block was incorporated into the architecture to process the new sequences while considering previously learned output. The memory block included several Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) memory cells. Furthermore, Luo et al. [96] developed a state estimation approach for power systems based on a broad learning method. The method can quickly compute the connecting weights among network layers and utilizes an IL system to update the modal based on new input data. The proposed approach was verified on a node system and with local power data. Similarly, Alves et al. [97] addressed the challenge of defending power system state estimation against false data injection attacks, which could manipulate input data such as measurements, system topology, or transmission line parameters. They proposed an IL Support Vector Machine (ILSVM) approach, which dynamically adjusted the SVM model in real-time through fast re-training to counteract cyber-attacks. Mollaiee et al. [98] proposed a novel CL scheme for online static security assessment to address uncertainties in modern power systems. Using a Mondrian forest-based model and a weather-dependent security index, the scheme periodically updates the assessment model.

Veerakumar et al. [99] proposed a Dynamic IL method for DL to address catastrophic forgetting in real-time disturbance event classification. The method, based on a replay-based IL strategy, minimizes training time, maximizes accuracy, and prevents forgetting. Zhu et al. [100] proposed a pipeline radial threat condition recognition model based on Multidimensional Information Fusion and a BLS (MIFBLS). The method improved signal processing, reduced false alarms, and enabled efficient real-time updates through IL. Experimental results on a natural gas pipeline dataset demonstrated the method's effectiveness in real-time monitoring and intelligent identification. Cai et al. [101] proposed an IL-enhanced LightGBM (IL-LightGBM) method for online load margin estimation in smart grids. By updating weight parameters with synchronized measurements, the method improves adaptability to operational variability. Case studies on IEEE 39-bus and 145-bus systems demonstrated their effectiveness and robustness in handling large-scale operational changes.

To conclude, CL significantly enhances fault and anomaly detection in the energy domain by enabling models to adapt to evolving data without retraining from scratch. As energy systems become increasingly complex, CL allows for the IL of new patterns, such as unexpected load spikes or equipment malfunctions, while retaining knowledge of previously detected anomalies. Techniques like IBL, replay, and EWC can prevent catastrophic forgetting, allowing the model to handle new fault types without interfering with previously learned knowledge. Future research could explore CL's application in improving the detection of novel and rare anomalies, addressing the challenges of class imbalance, and creating more adaptive and robust detection algorithms that can operate effectively in diverse and dynamic energy environments.

4.4. Load forecasting/prediction

Because of its relevance in energy management, infrastructure planning, and budgeting, electricity load forecasting has sparked interest in academia and industry. The spread of smart meters and other sensors in recent years has opened up new prospects for sensor-based load forecasting at the building and individual household levels. ML techniques for load forecasting have seen substantial success. However, these techniques use offline learning, which means they are taught just once and miss out on the potential to learn from freshly arriving data. Thus, deploying CL approaches solves that issue and provides new insights.

In [102], it was suggested to use EWC with a sliding window fine-tuning method for energy load prediction. The overview of the proposed approach is illustrated in Fig. 7. They evaluated this approach against three dynamic modeling techniques. The results indicated that the proposed method's mean absolute error decreased by 66.58%, 9.06%, and 8.70% on average compared to sliding window retraining, sliding window fine-tuning, and static modeling, respectively. Similarly, in their study, Li et al. [30] thoroughly examined the performance of three traditional model update techniques and five new CL methods using a 2-year dataset from 100 buildings sourced from an open database. The findings indicated that CL methods are more efficient at maintaining long-term accuracy while reducing computation time and data storage costs. Compared to static models and accumulative learning, the Coefficient of Variation of the Root Mean Squared Error (CV-RMSE) of EWC and GEM were around 14% and 8% lower, respectively.

In their study, Prabowo et al. [103] evaluated the effectiveness of using mobility data and a CL approach for forecasting buildings' electricity load. They employed Fast and Slow Network (FSNet) as the CL method with Temporal Convolutional Networks (TCN) as the backbone. The FSNet approach yielded good performance results with a lowest Mean Absolute Error (MAE) value of 5.26 for the pre-lockdown periods. Moreover, Fekri et al. [104] proposed an online adaptive RNN approach to continually forecast electrical loads. The RNN model captures time dependencies and updates weights according to new data to achieve CL functionality. The proposed approach was tested using data collected from five homes. Comparably, Hu et al. [105] utilized a reply approach for identifying the dominant load parameter. The backbone of the CL approach was a feed-forward neural network.

The authors in [106] showcased how catastrophic forgetting can affect forecasting performance in power grids. They evaluated various CL approaches against two scenarios where forgetting old knowledge can occur. The experimental work showed that EWC and Online-EWC yielded low forgetting values. In addition, Aragon and Chala [107] employed LSTM models for continuous load prediction. The evaluation and testing were conducted using residential datasets from various locations at different time intervals, such as hourly and minutely. The findings indicated that LSTM algorithms showed great potential for incorporating continuous load prediction with IL. That said, the paper in [108] proposed an adaptive approach for load forecasting using a combination of RNN and Autoregressive Integrated Moving Average (ARIMA) models. The suggested method's effectiveness has been tested on four households experiencing varying concept drift levels. The findings reveal that the combined approach outperforms the individual algorithms in terms of accuracy. Additionally, the study highlights the importance of evaluating load forecasting methods based on their ability to manage concept drift.

A framework called SteamDL was introduced in [109] to provide a DL interface with the Advanced Metering Infrastructures (AMI) data stream. The platform enabled the continuous learner to reduce the decline in performance resulting from changes in distribution. Moreover, the authors in [110] proposed a self-updating ML approach for building load forecasting. The ML model was based on the prophet model, which is, in turn, based on the General Additive Model (GAM). The self-updating segment of the system was based on using the experience replay CL approach. Similarly, Ramos et al. [111,112] employed an Artificial Neural Network (ANN) model with a CL approach to enhance load forecasting accuracy. The ANN model underwent daily retraining to ensure an up-to-date forecasting model. The study used the Weighted Absolute Percentage Error (WAPE) and Symmetric Mean Absolute Percentage Error (SMAPE) to evaluate the approach's effectiveness. Furthermore, the authors in [113] proposed an hourly continuous-learning system for load forecasting. The data was collected from various sensors, and then they were analyzed. Following that, a model is trained using that analyzed data if the model was not trained prior. If the model reappears, the model's performance will be evaluated first. If the performance is adequate, no training will be commenced; otherwise, the model will go through the retraining cycle, with the retrained model saved for future use. Additionally, Kim et al. [114] proposed an accelerated computing framework for accurate

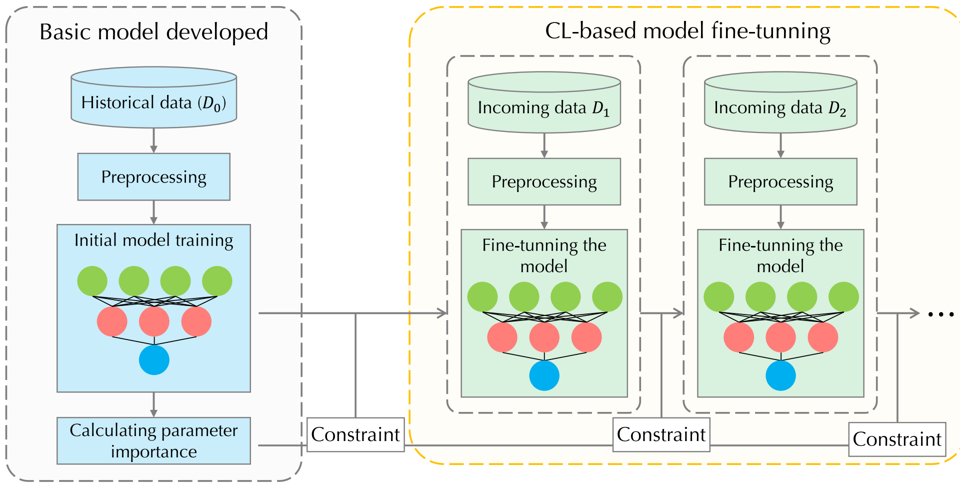

Fig. 7. Overview of CL-based load predicting approach [102].

energy prediction using an AMI data stream. The framework had two main elements: an adaptive incremental learner and a deep-learning accelerator using FPGA-GPU scheduling of resources. The experimental findings showed that the framework performed well for adjustable batch size and epoch for IL while ensuring minor forecasting errors, queue stability, an elevated model score, and cost-effective processing.

Moreover, the authors in [115] presented a novel approach to solving the concept drift and catastrophic forgetting issues by proposing an incremental solar power prediction approach. The approach trained a model to predict Photovoltaic (PV) values based on a BLS via a regional data exchange. Moreover, an online incremental model was proposed to learn the new features continuously and without forgetting. Similarly, the study described in [116] focused on predicting the hourly net load, which anticipates the disparity between the hourly power demand and the hourly power output of the PV system, representing the demand the utility needs to meet to the consumer. They primarily employed an LSTM and a Fully Online Sequential Extreme Learning Machine (FOS-ELM), an IL model that does not necessitate initial training data for this task. Also, Chen et al. [117] introduced an innovative lifelong learning approach using deep generative replay to dynamically and adaptively model building energy systems. Fig. 8 provides a schematic overview of the suggested lifelong learning technique for building energy systems. They employed a Variational AutoEncoder (VAE) to produce replay samples. To demonstrate the method's technical effectiveness, they conducted a field experiment in a custom net zero energy building to forecast solar power generation. The proposed method achieved an overall accuracy of 0.89, nearing the theoretical upper limit of 0.91 achieved through joint training.

On another note, Silva et al. [118] presented a smart grid-based load forecasting solution that utilizes a Fuzzy-ArtMap (FAM) and ANN. Training of the ANN involved using historical databases to extract fundamental knowledge. In addition to load forecasting, they implemented FAM-ANN CL for incremental knowledge extraction using real-time measurement system data. To validate the methodology, they utilized a historical database from an electric sector company, achieving a Mean Absolute Percent Error (MAPE) of approximately 5% without CL performance and generally less than 2% when CL execution was considered. Likewise, the authors in [119] presented EnGAT-BiLSTM, an enhanced framework for short-term load forecasting. The bidirectional LSTM aimed to support CL of load prediction. The proposed framework was designed to solve the data sparsity issue in short-term prediction and improve the overall prediction accuracy. Wang Ng et al. [120] proposed a novel DB-SOINN-R model for building load prediction, addressing the limitations of the Enhanced Self-Organizing Incremental Neural Network (ESOINN) model with density-based de-noising, a new distance metric, and k-nearest-neighbor with inverse distance weighting regression. The method outperformed five models in day-ahead and one-hour-ahead load predictions, demonstrating superior accuracy with IL on two datasets. Similarly, Chupong et al. [121,122] suggested using an OS-ELM model for hourly load forecasting, addressing the challenge of insufficient initial training data by synthesizing new samples with added noise.

Overall, CL enhances load forecasting in the energy sector by enabling models to adapt to changing demand patterns without forgetting previous knowledge. Approaches like replay-based methods, such as experience replay, allow the model to revisit past data, preserving important patterns while incorporating new information. Regularization techniques, such as EWC, help prevent catastrophic forgetting by penalizing large changes in model weights, ensuring stability as the model learns incrementally. Additionally, architectural approaches, like PNN, allow for the addition of new modules to handle new tasks without disrupting previous knowledge. Future research could focus on enhancing CL's ability to manage rare or extreme events, like sudden demand surges or grid disruptions, and scaling these approaches for large, distributed energy systems.

4.5. Renewable energy integration

Integrating renewable energy sources such as solar, wind, and hydro into energy systems helps to minimize reliance on fossil fuels. Energy management systems must optimize renewable energy generation, storage, delivery, and integration with the electrical grid.

Goh et al. [123] suggested using a Self-Organizing Neural Network (SONN) with an IL model to detect PV power fluctuation. The experiment result had a higher prediction rate of 95.83% compared to 93.81% in the simulation. Similarly, the authors in [124] introduced the usage of adaptive OS-ELM to detect fluctuations in a shipboard's electric power where renewable energy was integrated. The algorithm's efficacy is verified using real-time electric power fluctuation data from a ship under two distinct sea conditions.

The approach proposed in [125] used IL in conjunction with BLS to forecast solar irradiation with good performance in microgrid settings. The BLS demonstrated quicker learning capabilities than numerous DNNs due to its non-iterative nature and ability to bypass gradient descent while estimating the final output. Additionally, it is much easier to expand the network architecture using IL, allowing for the addition of enhancement nodes or mapped feature nodes to achieve the desired prediction and classification accuracy without retraining the network. The authors of [126] explored a new regression approach for solar irradiance detection using an incremental SONN model. The approach operates by incrementally learning the time-series solar irradiance and performing predictions in real-time. Almaraashi et al. [127] introduced

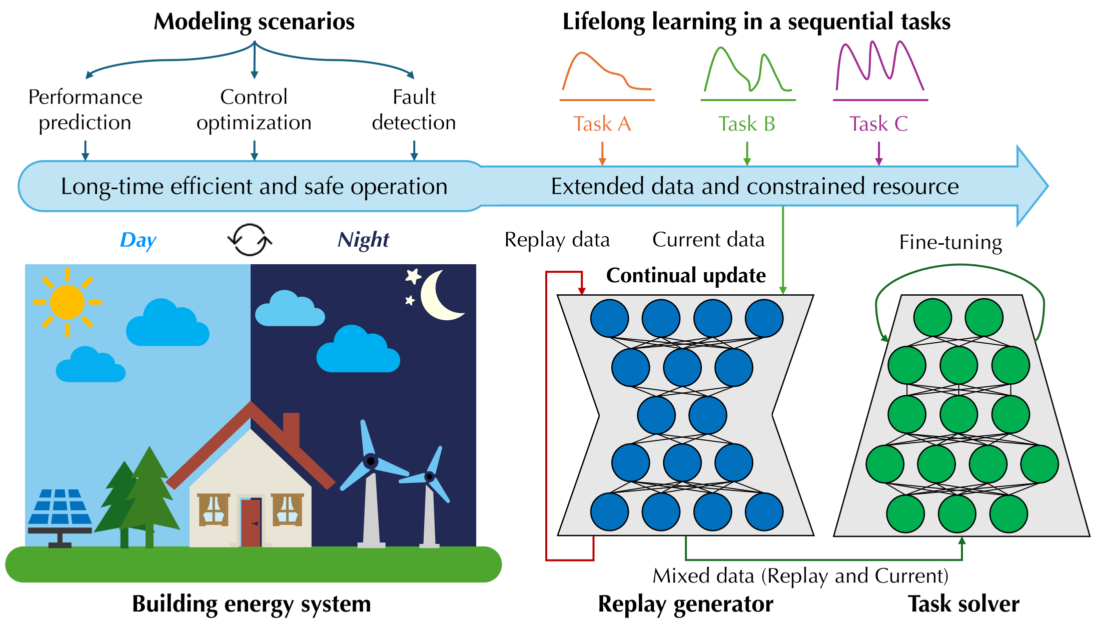

Fig. 8. Conceptual overview of performing CL in building energy systems [117].

a compact, explainable, lifelong learning-based interval type-2 fuzzy logic system for solar radiation modeling. Optimized using simulated annealing, the system produces accurate prediction models with many interpretable IF-Then rules. The lifelong learning approach ensures minimal forgetting while transferring model knowledge to new locations. The system outperformed other models by 13.2% and was well-received by experts for its transparency and potential for further model enhancement.

Sarmas et al. [128] employed a naive IL approach to load forecasting in microgrids. As for the DL, they utilized a Multi-Layer Perceptron (MLP). Experimental results showed that, in terms of the MAE metric, online learning models outperformed offline learning models by 8.6% in energy demand forecasting and 11.9% in renewable forecasting, emphasizing the benefits of IL. Furthermore, the work in [129] utilized a convolution technique to capture the spatial–temporal correlation between surrounding wind farms, and from this, a unique spatial–temporal wind power predictor, CSTWPP, was constructed. The proposed CSTWPP was trained in two steps, with initial offline training followed by gradual online training. Also, in the article [130], the authors suggested a framework for predicting targets and self-updating using CL. Experiments were conducted utilizing European wind farm data to assess the framework's performance in real-time. The experimental findings showed that the framework can learn from the data stream and increase prediction accuracy over time.

A unique hybrid architecture known as "CL-Net", based on Convolutional LSTM (ConvLSTM) and LSTM, was presented in the study in [131] for the multi-step forecasting of battery health and power consumption. Three datasets were utilized to confirm and assess the suggested architecture's efficacy. The comparison analysis demonstrated the efficacy and efficiency of the suggested design in multi-step battery state of health and power consumption forecasting. Comparably, the authors in [132] used CL and ML to assess battiers' performance. This type of battery is widely utilized in renewable energy storage because of its low cost, environmental friendliness, and scalability. Compared to the reference, the trained CL shows that it can assess the emergence of battery materials within the defined parameter space.

In a recent study, researchers proposed an innovative online risk assessment method for a power system characterized by high levels of renewable energy integration. This method was based on IL, as detailed in [133]. A case study was conducted using the IEEE-33 node system to validate the reliability of the risk assessment model. The study's results highlighted the effectiveness of the IL paradigm-based online risk assessment technique in predicting operational risk indicators. Furthermore, Yang et al. [134] investigated strategies for analyzing small signals in power networks with wind farm integration. A BLS model was built to assess the damping ratio sensitivity across several operational modes. The model was evaluated using the IEEE 3-machine, 9-bus, and New England 10-machine, 39-bus protocols. The findings show that the suggested strategy is feasible and effective. Also, the study in [135] offered a DRL technique for continually optimizing the control strategy of wind energy plants. The paper compared DRL to existing optimization methods such as Particle Swarm Optimization (PSO), Krill Herd, and Grey Wolf.

In [136], an OS-ELM-based technique is described to allow for fast real-time Dynamic Security Assessment (DSA) and model update. To improve the performance of ELMs, feature selection was performed using a single-feature estimate, and the findings were used to construct generation shifting as a preventative measure. The suggested approaches are evaluated using the New England 39-bus test system and compared to popular Intelligent Systems (IS) methods. The simulation results indicated that the ELM-based DSA approach has much faster calculation times while maintaining high, competitive accuracy. Lastly, Li et al. [137] proposed a system for secure energy management using a DRL algorithm and blockchain technology. The DRL algorithm dynamically adapts to varying demands and changing environmental characteristics. The system was evaluated through extensive simulations as well as real-world experiments.

Finally, CL approaches like replay, EWC, and LwF could seamlessly optimize the integration of renewable energy sources. This is done by enabling the models to adapt to changing production levels of renewable sources such as solar and wind. These CL strategies enhance the model's ability to forecast energy generation, optimize storage, and balance real-time supply and demand. Future research could improve CL's ability to handle extreme variability in renewable energy generation and integrate climate-specific data to enhance grid stability in dynamic environments.

Table 5 summarizes all the studies mentioned in this section where CL was applied in the energy domain based on several aspects, including the CL model(s) used, CL scenario, ML/DL approach used, application type, best performance value, and limitations.

Table 5
Comparison of studies using CL in the energy domain.

Ref.	CL approach	CL scenario	ML/DL approach	Application	Best performance value	Limitation
[71]	LwF and AIL	Class-IL	CRNN	Non-Intrusive Load Monitoring	Precision = 0.77, Recall = 0.78, and F1-score = 0.77	Additional factors that could impact performance were not explored
[72]	Replay	Domain-IL	LSTM and CNN	Non-Intrusive Load Monitoring	MAE = 5.18 and F1-score = 0.853	Benchmarking was not performed with CL baselines
[73]	AIL	Class-IL	ResNet-18	Non-Intrusive Load Monitoring	F1-score = 0.7069	Benchmarking was not performed with CL baselines
[74]	KD and Replay	Class-IL	CNN	Non-Intrusive Load Monitoring	F1-score = 0.7597	Benchmarking was not performed with CL baselines
[75]	Replay	Class-IL	ResNet-32	Non-Intrusive Load Monitoring	F1-score = 0.8195	Benchmarking was not performed with CL baselines
[76]	KD, Replay, and Weight Aligning	Class-IL	ResNet-20	Non-Intrusive Load Monitoring	Accuracy = 91.40%	Limited incorporation of unsupervised learning approaches
[78]	Kernel	General	MTL	Demand Response	N/A as numerical values	Benchmarking was not performed with CL baselines
[80]	MuZero RL	General	RF	Demand Response	MSE = 0.00603	The energy reduction ratio value was not determined
[81]	DRL	Domain-IL	LSTM	Peak Load Reduction	MAPE = 5.47%	The use of multi agent RL approach was not investigated
[82]	GEM	General	CNN-LSTM	Demand Prediction	ARMSE = 0.0905	The CosSim approach performed better than the study's proposed approach
[83]	Replacement Learning	Domain-IL	LightGBM	Electricity and Cooling Demand Detection	R2 = 0.969 and MAE = 0.004	Need to refine clustering techniques, improve the correction factor, and explore data characteristics
[84]	DRL	Domain-IL	LSTM	Demand-Aware Intelligent Pricing	N/A as numerical values	N/A
[85]	Replay	Task-IL	MLP	Energy Trading	N/A as numerical values	The framework's forgetting ratio was not measured
[86]	CWIB	Domain-IL	RBFNNs	Electricity Pricing Prediction	Accuracy = 75.80% and F1-score = 73.47%	Using different base models was not explored
[88]	OWM	Task-IL	CNN	Transient Stability Assessment	Test Accuracy = 97.48%	Performance metrics were limited to accuracy
[89]	KKT	Domain-IL	SVM	Transient Stability Assessment	Average Accuracy = 93.37%	Benchmarking was not performed with other CL approaches
[87]	SCP	Domain-IL	DRSN	Transient Stability Assessment	Accuracy = 98.03%	Number of scenarios was limited to three
[90]	IBL	Online	CNN	Voltage Stability Assessment	Average Accuracy = 98.38%	Other CL scenarios were not explored
[91]	IBL	Online	BLS	Transient Stability Assessment	RMSE = 0.296	An effective contingency filtering tool needs to be developed

(*continued on next page*)

Table 5 (continued).

[92]	Replay	Online	LightGBM	Transient Stability Assessment	Accuracy = 98.49%	Model's adaptability could be improved
[93]	KD and Replay	Online	Neural Networks	Transient Stability Assessment	Accuracy = 98.54%	Limited to simulation scenarios
[94]	GMM	Online	GMM and K-Means	Anomaly Detection in Smart Grids	Recall = 99.00%	Benchmarking was not performed with other CL approaches
[95]	LSTM and GRU	Task-IL	CNN and RNN	High Impedance Fault Detection	Accuracy = 99.48%	Generalization was not achieved
[96]	IBL	Online	BLS	State Estimation	Average relative error = 0.0111%	Benchmarking was not performed with CL baselines
[97]	ILSVM	Online	SVM	False Data Injection Detection	Accuracy = 97.40%	Benchmarking was not performed with CL baselines
[98]	MF	Online	MF	Security Assessment	Accuracy = 98.25%	Benchmarking was not performed with other CL approaches
[99]	Replay	Class-IL	CNN	Disturbance Event Classification	Accuracy = 97.30%	Gradual misclassification can occur
[100]	IBL	Online	BLS	Condition Recognition for Energy Pipeline	Accuracy = 90.12%	Research on spatial threat estimation of energy pipeline safety based on IoT systems is limited
[101]	IL-LightGBM	Online	LightGBM	Load Margin Estimation	RMSE = 0.59 and R2 = 0.973	Limited investigations into variable importance measures
[102]	EWC with a sliding window fine-tuning	Task-IL	ANN	Load Prediction	MAE = 293.183, RMSE = 452.487, CVRMSE = 0.200, and MAPE = 0.204	CL approach was not compared to other well known approaches
[30]	EWC, SI, LFL, MR, and GEM	Task-IL and Domain-IL	MLP	Building Energy Prediction	CV-RMSE of EWC and GEM decreased on average by 14% and 8% compared with static model and accumulative learning	Other DL models were not investigate
[103]	FSNet	Domain-IL	TCN	Electricity Load Forecasting	MAE = 5.26 for the pre-lockdown period	Benchmarking was not performed with other CL approaches
[104]	Online adaptive RNN	Online	RNN	Load Forecasting	MAE = 0.07 and MSE = 0.03 for 100-hours ahead	Benchmarking was not performed with other prominent CL approaches
[105]	Replay	Online	FFNN	Load Parameter Identification	N/A	Benchmarking was not performed with other prominent CL approaches
[106]	LWF, EWC, Online-EWC, and SI	Task-IL and Domain-IL	AE and MLP	Power Generation and Consumption Forecasting	MSE = 0.02049	Benchmarking was not performed with cumulative and joint training approaches
[107]	LSTM	Online	LSTM	Load Forecasting	MAE = 0.0044 and RMSE = 0.0048	Benchmarking was not performed with other prominent CL approaches
[108]	RNN and ARIMA	Online	LSTM	Load Forecasting	N/A as numerical values	Benchmarking was not performed with other prominent CL approaches

(continued on next page)

Table 5 (*continued*).

[109]	N/A	Domain-IL	DL	Individual Load Forecasting	N/A	Performance metrics were not mentioned
[110]	Replay	Domain-IL	GAM	Load Forecasting	MAPE = 13%	Performance metrics related to catastrophic forgetting were not explored
[111,112]	N/A	General	ANN	Load Forecasting	WAPE = 9.59%, SMPE = 13.64%	Performance metrics related to catastrophic forgetting were not explored
[113]	Naïve	Online	ANN	Load Forecasting	MAPE ≤ 2%	Performance metrics related to catastrophic forgetting were not explored
[114]	Adaptive IL	Domain-IL	CNN-LSTM	Energy Data Stream Processing	N/A as numerical values	Benchmarking was not performed with CL baselines
[115]	Ridge Regression	Online	BLS	PV Power Prediction	MAE = 0.012	Performance metrics related to catastrophic forgetting were not explored
[116]	FOS-ELM	Online	LSTM	PV Load Forecasting	MAPE = 1.9%	Performance metrics related to catastrophic forgetting were not explored
[117]	Generative reply with Conditional VAE	Task-IL	ANN	Solar Power Generation Forecasting	Accuracy = 89.00%	CL approach was not compared to other well known approaches
[118]	FAM	Online	FAM-ANN	Smart Grids Load Prediction	MAPE ≤ 2%	Performance metrics related to catastrophic forgetting were not explored
[119]	BiLSTM	Online	EnGAT-BiLSTM	Load Forecasting in Smart Grids	RMSE = 0.0423, F1 = 0.9412, and MAP = 0.9558	Complex scenarios were not investigated
[120]	DB-SOINN-R	Domain-IL	SOINN	Load Prediction	MAPE = 1.438%, RMSE = 0.979, CVRMSE = 2.071%, MAE = 0.667, and R2 = 0.996	N/A
[121,122]	FOS-ELM	Online	ELM	Short-Term Load Forecasting	MAPE = 1.36%	Benchmarking was not performed with CL baselines
[123]	TS-SOINN	General	SOM	Predict Power Fluctuation in PV	Prediction rate = 95.83%	The final network size was not optimized
[124]	AKOS	Online	ELM	Ship Power Fluctuations Prediction	ARMSE = 0.0188	Forgetting of older knowledge was not clearly explored
[125]	Architectural model	Online	BLS	Solar Irradiance Prediction	Depends on the testing scenario	Benchmarking was not performed with other CL approaches
[126]	E-SOINN	Online	RE-SOINN	Solar Irradiance Prediction	MASE = 0.65755, RMSE = 73.945	AI-based optimization algorithms were not explored
[127]	N/A	General	XAI	Solar Radiation Modeling	Average RMSE improvements of 13.2%	Limited by the need to refine pruning and consolidation processes

(*continued on next page*)

Table 5 (*continued*).

[128]	Naïve	Online	MLP	Photovoltaic Production and Load Forecasting	MAE = 6.697, RMSE = 13.260, and nRMSE = 0.527	Forgetting of older knowledge was not explored and benchmarking was not performed with other CL approaches
[129]	Naïve	Online	CNN	Wind Power Forecasting	MAE = 1.81, MASE = 71.6, and RMSE = 3.41	Challenges in model generalization across varying locations and seasons
[130]	EWC	Online	AE	Wind Power Forecasting	Prediction error = 0.623 and Forgetting ratio = 1.161	Comparative analysis was not performed with other CL approaches
[131]	ConvLSTM	Task-IL	ConvLSTM and LSTM	Batteries' State of Health and Power Consumption Forecasting	MSE = 0.015, RMSE = 0.122, and MAE = 0.088	Additional environmental parameters were not analyzed in the study
[132]	EWC and LwF	Task-IL	DNN	Battery Material Performance	Error <4%	Benchmarking was not performed with CL baselines
[133]	Replay	Online	LSTM	Risk Assessment for Systems using Renewable Sources	MAE = 0.095, MAPE = 2.012, MSE = 0.018, RMSE = 0.135, and RRMSE = 5.049	N/A
[134]	Naïve	Online	BLS	Stability Analysis of Power System Integrated with Wind Farms	MAPE = 0.17% and RMSE = 0.000106	Benchmarking was not performed with CL baselines
[135]	DRL	Online	DRL	Continuous Control Optimization in Wind Energy Systems	Overshoot = 0.0415% and settling time = 0.0011 secs	High computational complexity of DRL
[136]	Online Sequential	Online	ELM	Dynamic Security Assessment and Control	Accuracy = 98.5	Impact of missing data on the data-driven DSA methods was not investigated
[137]	DRL	General	DRL and Blockchain	Renewable Energy Management	N/A	Some scalability concerns

5. Case study for energy anomaly detection

Classification problems are the most common when it comes to ML algorithms and are used in most use cases where ML is employed. This is also the case for CL methods and applications; classification-based tasks are the most prevalent. In CL classification tasks, the goal is always to maintain good performance and eliminate the issue of catastrophic forgetting when introducing new input examples or classes. We implemented CL as a technique for detecting anomalies in power consumption. This approach is particularly focused on detecting Micro-Moments (MM), which are brief and significant fluctuations in energy usage that can indicate underlying issues or inefficiencies. By utilizing CL, we aim to enhance anomaly detection capabilities across various domains/tasks within the broader fault and anomaly detection context. MMs are derived from raw data collected by numerous sensors. An MM class is retrieved from a dataset using a rule-based model [138]. Table 6 summarizes the MMs feature classes derived using the proposed rule-based approach and their related label descriptions.

5.1. Dataset

5.1.1. Dataset description

Three different sets of data were used in this study to cover a wide range of situations and challenges. These datasets are the Dutch Residential Energy Dataset (DRED) [139], the Qatar University Dataset (QUD) [140], and SimDataset [140]. Each dataset offers distinct characteristics and challenges for the evaluation, ensuring a comprehensive analysis of CL model performance. The DRED dataset provides diversity in power contexts, the QUD dataset highlights power and environmental features, and the SimDataset introduces simulated scenarios. These three datasets allow for a detailed examination of the model's adaptability and generalization across various real-world and simulated conditions. The tabular data section of the study incorporates labels for the MM categories ranging from 0 to 4 in the DRED, QUD, and SimDataset. Table 6 elucidates the meaning of the MM labels.

5.1.2. Dataset pre-processing

An unbalanced dataset is an instance in which the distribution of samples across different classes is highly skewed, resulting in one or more classes having significantly fewer examples than others. This imbalance can lead to challenges in training ML models, as the model may become biased towards the majority class and have difficulty accurately predicting the minority class. Dealing with unbalanced datasets often requires techniques such as resampling, cost-sensitive learning, or different evaluation metrics to address the issue and improve model performance. The original distribution of the datasets was significantly unbalanced, as shown in Table 7. To address this issue, we decided to use the under-sampling strategy, which involves randomly removing instances from an over-represented class to balance the class distribution.

Table 6
Micro-moment labels and their descriptions.

Label	Case	Micro-moment description
0	Normal consumption	<95% of power consumption
1	Turn on	Switching on a device
2	Turn off	Switching off a device
3	Excessive consumption	>95% of power consumption
4	Consumption while outside	Device on but user is not present

Table 7
Distribution of micro-moment labels across the three datasets.

Dataset	Label					Total
	0	1	2	3	4	
DRED	45,455 (27%)	3,315 (2%)	3,342 (2%)	35,044 (21%)	79,196 (48%)	166,352
QUD	12,102 (26%)	1,568 (3%)	1,580 (3%)	3,954 (8%)	27,725 (59%)	46,929
SimDataset	59,424 (57%)	7,780 (7%)	7,779 (7%)	6,343 (6%)	23,793 (23%)	105,119

Moreover, most open-source CL approaches are based on using images as inputs. The MM tabular data was transformed into images using the approach found in [141]. The case study closely aligns with the Domain-IL CL scenario, as the three datasets were collected in different contexts and domains, resulting in varying data distributions; however, they share the same labels/classes. Additionally, a data augmentation technique was incorporated to further evaluate the effectiveness of the CL approach by introducing various noise variations to the original data. This allowed us to create an additional task for each dataset, resulting in a total of six tasks for both training and testing purposes. So overall, we will use six tasks, therefore six different MM-based datasets, sharing the same number of five classes.

5.2. Experimental setup

5.2.1. Base model

We have utilized ResNet-18 as the base model, which comprises 18 layers and functions as a CNN model. In Python, we have employed a pre-trained version of the network trained on more than a million images from the ImageNet database [142]. The pre-trained network can classify images into 1000 categories, such as keyboards, mice, pencils, and animals. As a result, the network has developed advanced feature representations for a wide range of images. The network takes images with a size of 224-by-224 as input. We have modified the output layer to specifically classify only five categories instead of the original 1000 to customize the network for our specific requirements.

5.2.2. Employed CL methods

In terms of the *replay methods*, we have utilized the basic rehearsal [40], Gradient Episodic Memory (GEM) [44], and Average GEM (AGEM) [45]. As for the *regularization-based methods* we deployed the Elastic Weight Consolation (EWC) [143], Synaptic Intelligence (SI) [144], Memory Aware Synapses (MAS) [145], and LwF [38]. Finally, for the *architectural-based approach*, we have utilized the Copy Weight with Reinitialization Star (CWR*) [146] method. For our implementation, we have utilized the Avalanche library [147,148] for CL. Avalanche, developed in PyTorch, is a CL Library designed for End-to-End functionality. It was created in ContinualAI to offer a collaborative open-source platform for rapid prototyping, training, and reproducible assessment of CL algorithms.

5.2.3. Evaluation metrics

In the context of CL, we assess performance by measuring the average accuracy and forgetting of each task after its re-training. Forgetting is the difference between the initial task knowledge, represented by the accuracy during the first learning phase, and the accuracy achieved after training on one or more additional tasks. In the visual representations, we focus on tracking the accuracy changes for each task as more tasks are introduced. The legend presents the mean accuracy and mean forgetting for the final model, which is determined by assessing each task after the entire learning sequence.

5.2.4. Baselines

The employed CL methods were meticulously compared against various baselines:

1. Naive begins by optimizing the prior task model to the present task's parameters. This baseline greedily trains each task without considering past task performance, resulting in catastrophic forgetting and indicating the lowest required performance.
2. Joint training simultaneously evaluates all data in the task sequence, violating the CL paradigm. This baseline establishes a goal reference performance.
3. Cumulative training involves training the current task based on the knowledge acquired from the previous task while incorporating the earlier tasks' training samples. Just like joint training, this approach achieves optimal performance.

5.2.5. Learning attributes

Throughout the training process, the models underwent optimization using stochastic gradient descent with a momentum value of 0.9, and the learning rate was specifically set to 0.001. The training phase was carried out for ten epochs for each task. Additionally, a training-to-testing ratio of 80/20 was chosen to ensure comprehensive evaluation and validation of the model's performance.

5.3. Results and discussion

The outcomes of implementing the different CL methods can be seen in Fig. 9 alongside the baseline results. Fig. 9 displays the outcomes for all three CL families. Each figure comprises six subpanels, each subpanel illustrating the evolution of test accuracy for a specific task (e.g., Task 1 for the leftmost panel) as additional tasks are included in the training. As the nth task is included in the training only after n steps, the curves become shorter as we progress to subpanels on the right.

The naive approach showed the poorest performance due to severe catastrophic forgetting. It initially achieved good results when training for a task, but when re-training for a new task, the performance of the old task quickly degraded. This resulted in an average accuracy of 26.35% and a forgetting rate of 52.42%. On the other hand, the joint training and cumulative methods produced the best results, with an average accuracy of at least 86.00%.

Regarding replay-based approaches, the basic rehearsal approach demonstrated the highest average accuracy performance at 70.26% and a forgetting rate of 9.47%. This was accomplished with a memory buffer of 200 samples for each task. The GEM and AGEM approaches displayed similar performance, with an average accuracy of around 50.00% and a forgetting rate of 16.00–18.00%, respectively.

Among the various regularization-based approaches utilized in the case study, the MAS performed the best in terms of average accuracy

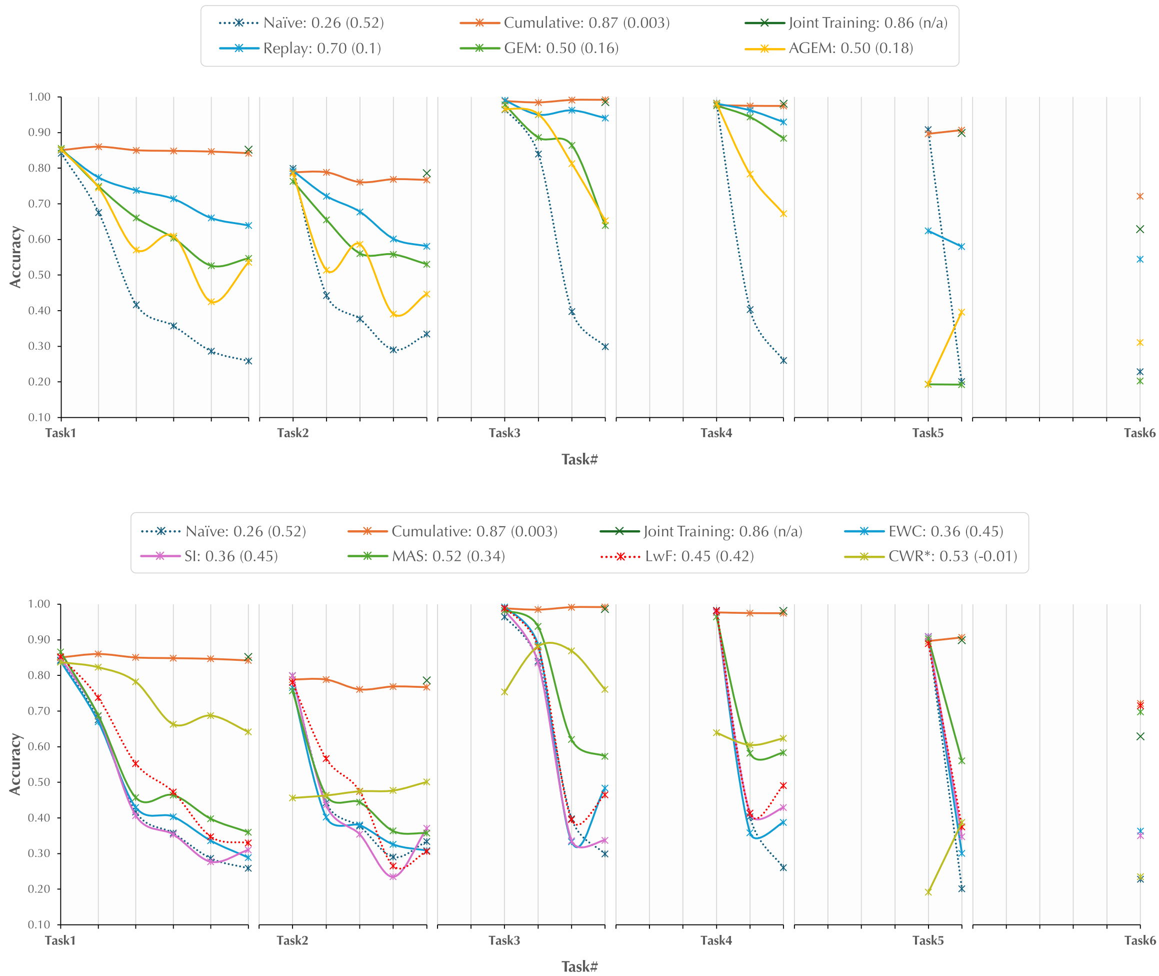

Fig. 9. Replay methods (top) and regularization-based and architectural methods (bottom) on ResNet-18 for the base model, reporting average accuracy (forgetting) in the legend.

and forgetting rate of 52.19% and 33.96%, respectively. The second best was the LwF, which reported an average accuracy of 44.68% and a forgetting rate of 42.10%. The EWC and SI performed similarly across the two metrics. The CWR* reported the lowest average forgetting rate of −1%. Additionally, comparing the CWR* approach with other regularization approaches, it achieved the highest average accuracy of 52.50%. CWR isolates task-specific weights and carefully manages how new and old knowledge is merged, preventing new tasks from interfering with previously learned tasks.

Future work in this context could involve conducting the anomaly detection case study in a semi-supervised or even unsupervised manner, providing valuable insights into the robustness of the CL methodologies. In a semi-supervised approach, a small amount of labeled data could be combined with a larger pool of unlabeled data. On the other hand, exploring an unsupervised approach would entail analyzing data without any prior labels, allowing the system to discover patterns and relationships on its own. Both approaches present unique advantages and challenges, and investigating them could lead to a deeper understanding of the CL methodologies employed in the study.

6. Key challenges

6.1. Stability–plasticity dilemma

The stability–plasticity dilemma is a fundamental challenge in CL, particularly in the energy domain and in tasks like energy consumption prediction, where new devices, sensors, and consumer behavior patterns constitute a continuously evolving landscape. These dynamics raise the need for a smart grid system that can adapt to the new variables (*plasticity*) while retaining the previously acquired knowledge (*stability*) [83]. The same dilemma arises in the case of fault detection in energy networks and their components (e.g., energy cells, power lines, etc.) when new smart devices, sensors, and metering systems are incorporated and consequently introduce new types of failures [149]. Achieving an optimal balance between the two is essential for successful CL systems [30].

6.1.1. Stability and plasticity

Stability refers to a model's capacity to maintain its performance on previously learned tasks when faced with new data or tasks. In the context of CL in energy systems, stability is crucial to avoid the well-known phenomenon of *catastrophic forgetting*, where the model loses its ability to recall previously learned information after being exposed to new energy consumption patterns or renewable energy generation forecasts [150]. For example, in the case of energy price forecasting, the models have to be continuously trained to predict price fluctuations due to technological advancements (e.g., new energy storage systems), supply chain disruptions (e.g., geopolitical crises) or policy changes (e.g., new taxes) and at the same time must retain important historical patterns to maintain accurate forecasting [83].

6.1.2. Trade-off and the dilemma

The stability–plasticity dilemma illustrates that a balance must be struck between these two competing objectives [151]:

- **Too much stability** results in a rigid model that is resistant to change and cannot adapt well to new tasks. While it may retain previously learned knowledge effectively, the model's ability to learn new information is impaired. This can lead to poor performance on newer tasks as the model becomes overly biased towards previously seen tasks [152].
- **Too much plasticity** makes the model overly adaptive to new tasks, potentially forgetting previously learned knowledge. While plasticity can be beneficial for adapting to new trends, excessive plasticity can lead to overwriting knowledge from earlier tasks [153]. Techniques such as experience replay, regularization, and incremental and MTL can improve the balance between historic and new knowledge [154].

6.1.3. Mathematical perspective

From a mathematical perspective, the trade-off between stability and plasticity can be framed in terms of optimizing model parameters. In standard neural networks, weights are adjusted during training via gradient descent [155]. When learning new energy tasks, gradient updates may lead to significant shifts in model weights, particularly if the new task requires significantly different representations from previous tasks, such as adjusting to a new energy source or demand pattern. This shift is the essence of plasticity, but without mechanisms to preserve important weights for prior tasks, the model's stability is compromised [156].

The *stability–plasticity dilemma* can thus be framed as a constrained optimization problem, where the objective is to minimize the loss on the current task while ensuring minimal disruption to the model's performance on previously learned tasks. This requires approaches that explicitly manage how model parameters are updated, preventing critical weights from being overwritten while allowing enough flexibility to learn new tasks [157].

6.1.4. Challenges in achieving balance

Balancing stability and plasticity is difficult due to several factors:

- **Non-Stationary Data**: In real-world CL scenarios, data distributions change over time. A model must adapt to new data distributions (plasticity) without losing performance on previously learned data (stability). This becomes challenging when new tasks differ significantly from old ones [158]. For example, energy demand can vary drastically from winter to summer, driven by heating or cooling needs. These fluctuations, which are due to seasonal changes, create non-stationary demand data over time [159].
- **Task Similarity**: When new tasks are similar to old ones, it is easier to strike a balance between stability and plasticity because the representations learned for the new tasks can be built upon the old. However, when tasks are dissimilar, a high degree of plasticity is required to learn the new tasks, increasing the risk of catastrophic forgetting [160]. For example, a model trained to predict energy prices and later tasked with forecasting energy demand will face stability problems because the two tasks may be similar but not identical. Both tasks depend on factors like weather, consumer behavior, and supply, but the relationships may differ. Similarly, the tasks of short-term and long-term prediction of energy prices differ since they should consider different factors (e.g., weather conditions and supply disruptions versus policy and market trends).
- **Capacity Limitations**: Neural networks have finite capacity, meaning they can only store a limited amount of knowledge. As the number of tasks increases, it becomes harder for the model to maintain stability while learning new tasks, as the new information competes with old information for representation in the network [66]. The limitations in the energy domain can be reached, for example, when an energy forecasting model is scaled up to cover multiple regions, which involves handling large datasets with diverse geographical and temporal variations. If the model is retrained frequently with new data from one region, it might forget the consumption behaviors of other regions it was previously trained on, affecting the model's generalization ability [161].

6.1.5. Existing approaches to address the stability–plasticity dilemma

Several techniques have been proposed to manage the trade-off between stability and plasticity in CL. Regularization-based approaches address the stability–plasticity dilemma by preserving critical parameters of previous tasks through regularization penalties. Methods such as EWC [61,162], SI [144,163], and MAS [164] reduce catastrophic forgetting by constraining changes to weights deemed important for prior knowledge, thus maintaining stability without hindering new learning.

Replay strategies mitigate forgetting by revisiting prior knowledge during training, thus ensuring the balance between retaining stability and enabling plasticity. Experience Replay [165] achieves this by interleaving stored examples from past tasks, while Generative Replay [166] synthesizes pseudo-data resembling earlier tasks. Conversely, architectural solutions dynamically adjust the model structure to accommodate new tasks while safeguarding prior knowledge. PNNs and Dynamic Parameter Isolation [167] either expand the architecture or selectively isolate task-specific parameters, preventing interference with earlier learning. While these approaches excel in preserving stability, scalability challenges can arise with an increasing number of tasks.

6.2. Data scarcity

Data scarcity is an important challenge in CL, which is due to the limited availability of sufficient and diverse data for each new task. Usually, CL systems must learn from limited, sometimes imbalanced data, leading to challenges in generalization and adaptation [168]. Effective CL requires strategies to overcome these limitations while maintaining the model's ability to adapt to new tasks and retain past knowledge.

6.2.1. The impact of limited data

Data scarcity can lead to two major risks: overfitting and underfitting. Overfitting occurs when the model becomes overly tailored to the small dataset of the current task, losing its generalization capabilities [169]. This makes the model less flexible (reduced plasticity), as it is biased towards the specific instances it has seen. On the other hand, underfitting happens when the model fails to capture the essential patterns in the data due to its limited size or diversity, resulting in poor performance on the new task. This is the case of energy forecasting models in buildings that are equipped with smart meters with varying data sampling granularities or in which there is a lack of historical data [170]. The volatility of the population and the continuously changing user profiles. At the same time, the model may suffer from catastrophic forgetting if it cannot adequately consolidate knowledge from previous tasks due to insufficient exposure to new data [171]. Thus, it becomes harder to achieve robust performance on old and new tasks with limited information.

6.2.2. Challenges in task representation

Data scarcity often complicates the representation learning process, particularly when tasks are highly varied. In multi-task setups [172], the model must handle a sequence of tasks and relies on data to develop meaningful representations that generalize across tasks [173]. When data is scarce, the model may fail to develop these representations adequately, resulting in suboptimal performance on both the current and previous tasks. Furthermore, the absence of enough data from certain tasks can lead to biased representations, affecting the model's long-term stability and plasticity.

6.2.3. Existing approaches to mitigate data scarcity

Several methods have been proposed to alleviate the challenges posed by data scarcity in CL. First, data augmentation techniques, such as rotations, scaling, and cropping, artificially have been used to increase the diversity of training data, improving generalization and reducing overfitting [174]. Generative models, such as Generative Adversarial Networks (GANs) and VAEs, offer another solution by creating synthetic data that mimic the distribution of scarce datasets, effectively augmenting the training set [175]. Additionally, self-supervised learning approaches enable the use of unlabeled data to pre-train models, which can then be fine-tuned on scarce labeled data, demonstrating robustness in various applications, including fault diagnosis in HVAC systems [176,177].

Transfer learning and KD have also been introduced as promising strategies to counteract data scarcity. Transfer learning allows models to leverage knowledge from domains with abundant data to improve performance in target domains with limited examples, as demonstrated in energy management tasks like power consumption forecasting [178]. This approach ensures stability in pre-trained knowledge while enabling flexibility to adapt to specific new tasks. Similarly, KD transfers essential information from a larger, well-trained model to a smaller, more efficient one, making it ideal for tasks with constrained data, such as electric load forecasting in individual households [179]. These techniques balance stability and plasticity, ensuring adaptability without compromising the retention of prior knowledge.

Finally, replay methods, meta-learning, and task-specific architectures have been suggested to mitigate data scarcity. Experience replay allows for enhanced stability while supporting adaptation [165] to new tasks, whereas meta-learning techniques, including few-shot learning, maintain plasticity without sacrificing stability [180]. Task-specific architectures, such as modular networks, assign dedicated sub-networks to different tasks, enabling specialization without requiring extensive new data [181]. All the aforementioned methods collectively mitigate the challenges posed by limited data, supporting the effective application of CL in diverse scenarios.

6.3. Security and privacy

Security and privacy are critical challenges in CL, particularly when models operate on sensitive data or are deployed in environments where they are exposed to adversarial attacks. Adversarial Learning Attacks (ALAs), which involve tampering with meteorological data from external Application Programming Interfaces (APIs) to disrupt renewable energy forecasting, can compromise CL models by embedding malicious patterns, leading to inaccurate predictions and jeopardizing power system operations [182]. As CL systems are designed to interact with evolving data streams, ensuring that these systems are protected from malicious actors and that user data remains confidential is paramount for their safe deployment and operation. Maintaining a balance between learning new tasks and ensuring robust security and privacy protections is a significant challenge in CL. Implementing advanced intrusion detection systems that leverage ML is essential for identifying cyber threats in CL models. These systems must be continually updated with diverse attack scenarios to adapt to evolving threats and provide timely alerts for effective defense. Implementing advanced intrusion detection systems that leverage ML is essential for identifying cyber threats in energy systems. These systems must be continually updated with diverse attack scenarios to adapt to evolving threats and provide timely alerts, ensuring the resilience and security of energy forecasting and management [183].

6.3.1. Privacy concerns

In CL, models often require access to diverse and potentially sensitive datasets across multiple domains. Without proper privacy mechanisms, user information might be leaked or exposed during training, especially when learning new tasks from decentralized sources or when training data is distributed across various devices [184]. Techniques like FL have been proposed to enhance privacy in energy management applications by keeping data localized and only sharing model updates. This approach is particularly useful for tasks like net-energy forecasting, where sensitive household or industrial data must remain secure. However, even with such techniques, vulnerabilities such as inference attacks remain a concern, as attackers can still potentially extract sensitive information from the model's updates, compromising privacy despite the system's distributed nature [185].

6.3.2. Security threats

CL systems are susceptible to various security threats, including adversarial attacks, data poisoning, and model inversion attacks. Adversarial attacks attempt to manipulate the model's learning process by introducing small perturbations to the input data, which can lead to incorrect predictions or significant performance degradation. Data poisoning is another critical threat, where an attacker intentionally introduces misleading data during the learning process to corrupt the model [186]. This is especially dangerous in CL, as the model's ability to adapt may make it more prone to learning from poisoned data.

6.3.3. Existing approaches to address security and privacy

- **Differential Privacy**: Differential Privacy (DP) is a widely adopted technique that safeguards user privacy by ensuring that the inclusion or exclusion of any individual data point minimally impacts the model's output. In energy grids, DP can be applied by adding noise to energy consumption data used for demand forecasting, preventing the leakage of sensitive household patterns [187]. Integrating DP into CL systems involves injecting noise into model updates and balancing privacy protection with model performance, as excessive noise can impair learning accuracy.
- **Adversarial Training**: Adversarial training involves augmenting the training data with adversarial examples to improve the model's robustness against attacks. For instance, an ensemble adversarial training-based robust model can be employed for multi-horizon dynamic line rating forecasting, enhancing resilience against adversarial attacks [188]. This approach can also be applied to CL systems to strengthen their defenses while learning new tasks [189]. However, adversarial training often introduces computational overhead and requires careful tuning to balance security with learning efficiency.
- **Federated Learning and Secure Aggregation**: FL enables decentralized model training, where data remains on local devices, and only model updates are shared. This helps preserve user privacy by preventing raw data from being transmitted to a central server [185]. To further enhance privacy, secure aggregation techniques can be applied to ensure that individual device updates are encrypted before being aggregated, reducing the risk of information leakage during the learning process.

6.4. Scalability and interoperability

Scalability and interoperability represent significant challenges in CL, particularly as the complexity and size of models and datasets increase. As CL systems are deployed in diverse environments and must interact with various external systems, ensuring they can scale efficiently while maintaining seamless interoperability is crucial for their effectiveness and applicability in real-world scenarios [172].

6.4.1. Scalability challenges

Scalability refers to a system's ability to handle increased loads, whether in data, model complexity, or the number of tasks learned. In the energy domain, this could involve forecasting electricity demand across a growing number of regions, processing vast datasets from smart meters, or integrating diverse renewable energy sources into grid management systems. As the number of tasks increases, models must be able to efficiently process and learn from large volumes of data without significant degradation in performance or increase in computational resource demands [190]. Additionally, training on larger datasets, such as historical energy consumption data or meteorological inputs for renewable energy forecasting, can lead to longer training times, making it essential for CL systems to employ strategies that ensure rapid adaptation and training efficiency, such as distributed learning frameworks or online learning methods.

6.4.2. Interoperability concerns

Interoperability refers to the ability of different systems and models to work together and share information effectively. In the energy domain, this is particularly important when models trained on diverse datasets, such as smart grid data, renewable energy forecasts, or energy consumption patterns, need to collaborate or integrate insights from one another. Challenges arise due to differences in model architectures, data formats, and training protocols, which can hinder effective communication between systems. For example, integrating energy forecasting models with grid management systems or combining data from various sensors and meters with centralized platforms can create significant interoperability issues. Ensuring that models can operate across different platforms and with various types of data requires the establishment of standardized protocols and frameworks [191].

6.4.3. Balancing scalability and interoperability

Achieving a balance between scalability and interoperability is essential for deploying CL systems in practical applications. While scalable systems can efficiently handle large amounts of data and tasks, they must also be designed with interoperability in mind to facilitate collaboration with other models or systems. This balance can be challenging, as increased complexity in scaling systems may lead to greater difficulty in ensuring that they can communicate effectively with other models [192].

6.4.4. Existing approaches to address scalability and interoperability

- **Distributed Learning Frameworks**: Distributed learning frameworks enable the training of models across multiple nodes, which can significantly improve scalability [193]. By parallelizing the learning process, these frameworks allow for faster processing of large datasets and the ability to manage numerous tasks concurrently without overwhelming a single system. However, ensuring interoperability among distributed systems remains challenging, as data and model updates must be synchronized effectively.
- **Standardized Protocols**: Implementing standardized protocols for data representation and communication can enhance interoperability in CL systems. Techniques such as the Open Neural Network Exchange (ONNX) format provide a common framework for sharing models across different platforms and frameworks, allowing seamless integration of diverse learning systems [194]. Standardization can help facilitate the exchange of information between models trained on different tasks or datasets.
- **Modular Architectures**: Modular architectures enable components of CL systems to be developed and deployed independently, allowing for greater flexibility and scalability. For example, in energy management systems, different renewable energy forecasting, demand response, and grid optimization modules can be independently updated or replaced without disrupting the entire system. By designing systems with interchangeable modules, it becomes easier to adapt to new tasks or incorporate new technologies, such as AMI or smart grid technologies without disrupting existing functionality. This modular approach also supports interoperability by allowing different components to communicate through well-defined interfaces.

6.5. Expanding the practical applications of CL in energy systems

The practical deployment of CL in energy systems faces several critical challenges. Energy systems demand real-time decision-making, requiring research to optimize CL methods for minimal latency while effectively adapting to dynamic data streams [114]. Additionally, many energy systems operate under computational and resource constraints, such as edge devices, necessitating the development of lightweight CL frameworks and efficient task allocation strategies to enhance operational efficiency [119,120]. Interoperability across diverse energy infrastructures and platforms remains another significant hurdle, emphasizing the need for standardized protocols to seamlessly integrate CL models [110,140]. Moreover, adopting AI-driven CL systems must address regulatory and policy considerations, particularly regarding data privacy, security, and fairness, to ensure compliance and public trust in energy management applications [124,137].

6.5.1. Smart grid management

Smart grids benefit significantly from CL applications such as dynamic load forecasting, energy demand-response optimization, real-time anomaly detection, and fault diagnosis [114,118]. However, integrating CL into existing grid infrastructure presents challenges, including managing data heterogeneity across regions and ensuring system reliability amid dynamic learning updates [119,120]. Future research should focus on developing task-specific CL models tailored for grid operations, establishing real-time performance benchmarks, and optimizing communication networks to enhance the efficiency of distributed CL systems in smart grid environments [113,123].

6.5.2. Renewable energy integration

CL has promising applications in renewable energy, including predicting energy generation from sources like solar and wind, optimizing energy storage systems, and enhancing grid stability amidst fluctuating inputs [117,130]. However, significant challenges remain, such as managing concept drift caused by seasonal variations, training models on incomplete or sparse datasets, and adapting to rapid technological advancements in renewable energy systems [115,116]. Future research should prioritize hybrid CL approaches, combining replay methods and architectural solutions to address fluctuating renewable outputs, integrating climate-specific data for enhanced adaptability, and improving scalability to efficiently manage larger distributed energy systems [124,125].

6.5.3. Building energy management

CL offers transformative applications in smart buildings, including personalized energy consumption forecasting, anomaly detection in energy usage, and optimizing HVAC systems [110,140]. However, challenges persist, such as safeguarding privacy in residential energy data, addressing the limited generalizability of models across diverse building types, and seamlessly integrating CL with IoT devices [30,104]. Future research should focus on implementing privacy-preserving CL methods like FL, developing modular architectures for building-specific adaptation, and deploying lifelong learning systems that continuously adapt to evolving conditions over extended periods [103,105].

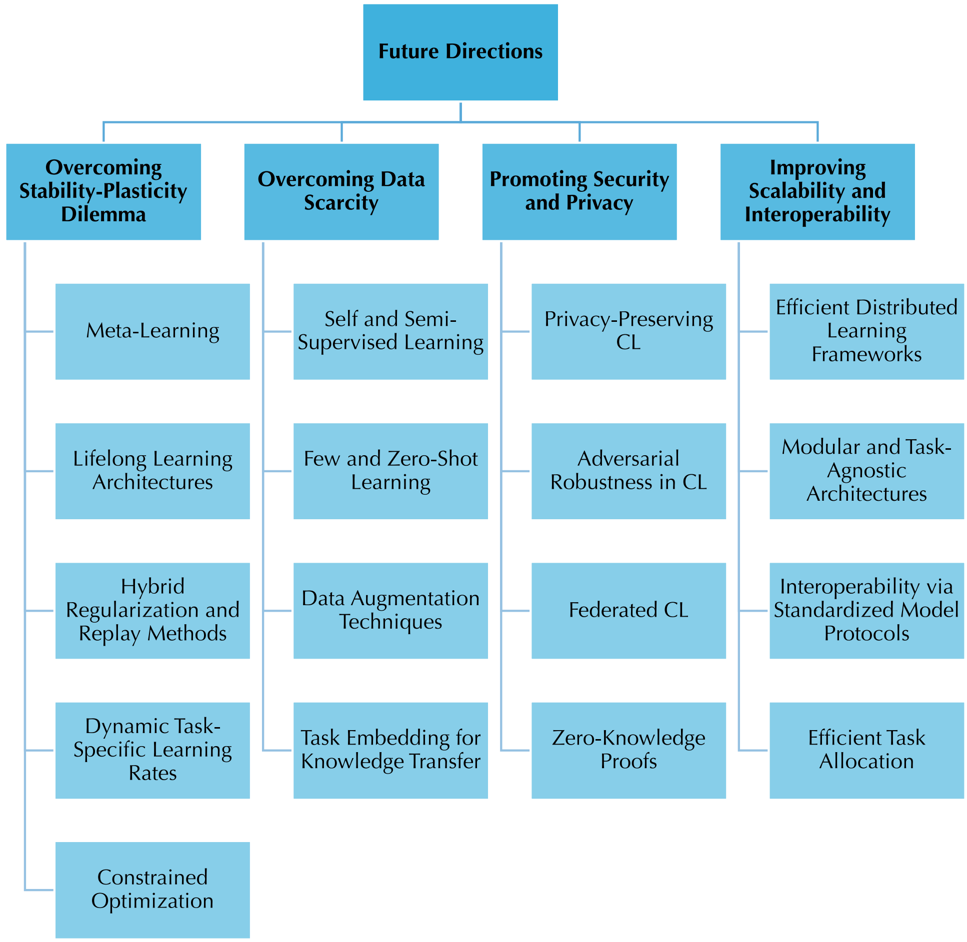

Fig. 10. Future directions that could address the challenges associated with CL deployment.

6.5.4. Energy trading and market optimization

CL has significant applications in dynamic pricing models, real-time prediction of energy demand-supply imbalances, and adaptive energy trading strategies [119,137]. However, key challenges include mitigating adversarial threats in energy trading, maintaining fairness in pricing algorithms, and ensuring interoperability within decentralized energy markets [106,123]. Future research should focus on developing adversarially robust CL models for secure energy trading, leveraging Zero-Knowledge Proof (ZKP) methods for secure data sharing, and integrating blockchain technology to ensure tamper-proof transactions and enhance system reliability [124,137].

7. Future directions

Addressing the key challenges of CL, such as the stability–plasticity dilemma, data scarcity, scalability, and security requires advancing research in several areas. Below are future research directions, and advancements from the broader area of ML/DL that can help overcome these challenges in the energy domain (see Fig. 10).

7.1. Overcoming stability–plasticity dilemma

7.1.1. Meta-learning approaches

Meta-learning has emerged as a powerful approach to improve CL by addressing critical challenges such as catastrophic forgetting and the stability–plasticity dilemma. Traditional methods often struggle with the balance between adapting to new tasks while retaining knowledge from previous one, but meta-learning techniques offer solutions by optimizing how models learn and update their parameters. One of the key contributions to this field is the Look-ahead MAML (La-MAML) algorithm, proposed by Gupta et al. [195], which introduces a fast, optimization-based meta-learning method tailored for online CL. The La-MAML algorithm can be adapted to energy load forecasting by modulating per-parameter learning rates during meta-learning updates, enabling the model to adapt efficiently to dynamic load patterns. Additionally, its use of a small episodic memory to store and replay key historical energy data can help maintain forecasting accuracy while mitigating catastrophic forgetting.

In another study, Javed and White [196] focused on reducing interference between tasks while enhancing fast adaptation to new ones. Their method, an extension of the MAML framework, emphasizes minimizing task interference and improving task adaptation by learning representations that support both objectives. By incorporating these features, the model performs better on new tasks while retaining knowledge from previous tasks, significantly reducing the forgetting issue often observed in CL. Addressing the stability–plasticity dilemma, Han et al. [197] introduced a framework that optimizes CL through a meta-learning-based approach. Their method, MMKDDA, uses multi-scale KD to manage the long-range and short-range spatial relationships between features, which helps mitigate the data imbalance between tasks. Blending data from past and current tasks during training helps maintain stability for past knowledge while providing the flexibility needed for learning new tasks. The meta-learning update in their framework balances these conflicting demands, thus improving stability and plasticity in CL. Finally, Son et al. [198] explored how combining meta-learning and CL frameworks can address the complexities of streaming data. Meta-learning, often described as “learning to learn”, optimizes the learning process, making it well-suited for environments where tasks arrive sequentially and must be learned incrementally. By continually adapting the learning algorithm, models can better manage new tasks while retaining performance on older ones. This synergy between meta-learning and CL opens new avenues for research and innovation, providing a framework for developing models that excel in dynamic environments. Such approaches highlight the value of meta-learning in fine-tuning the balance between retention and adaptability in sequential learning settings and seem appropriate for the energy domain prediction tasks.

7.1.2. Lifelong learning architectures

Research on neural networks with adaptive architectures that grow in response to new tasks without compromising old knowledge could provide a solution. PNNs or modular architectures that selectively update parts of the model based on task similarity might offer a better balance between stability and plasticity. For instance, in next-generation green building energy systems, novel lifelong learning methods such as deep generative replay have been proposed for dynamic and adaptive modeling to enhance energy prediction, predictive maintenance, and control optimization. These methods alternate training between a task solver and a replay generator, allowing models to retain previous knowledge while adapting to new tasks without explicitly storing data, thereby conserving resources and protecting privacy. Typically, lifelong learning architectures enhance CL by providing models with the ability to continuously adapt to new data, retain past knowledge, and improve performance over time. Whether in anomaly detection or dialog systems, lifelong learning enables models to handle evolving patterns and dynamic environments, making them more robust and scalable while reducing the need for manual intervention [199,200]. In the case of solar power generation prediction for net-zero energy buildings, such methods achieved a 53.4% higher accuracy than standard approaches and effectively reduced forgetting rates to below 0.10, demonstrating their feasibility and superiority over traditional retraining-based methods [117].

7.1.3. Hybrid regularization and replay methods

This hybrid method can significantly improve CL by mitigating catastrophic forgetting and enhancing the model's ability to retain and integrate knowledge from previous tasks while learning new ones. Replay methods, such as those discussed by Merlin et al. [201], work by storing a subset of past data, which is then used during training to reinforce prior knowledge, helping to balance memory efficiency and performance. These replay strategies, combined with data augmentation, further improve model robustness even with smaller memory buffers. Additionally, Han et al. [202] highlighted the effectiveness of hybrid approaches that combine regularization and replay techniques. Regularization strategies can stabilize learned representations, preventing model outputs from drastically shifting when learning new tasks. This dual regularization, particularly in domain-IL, ensures that the model retains past knowledge while adapting to new environments. Lomonaco et al. [203] extended this idea by proposing a more flexible hybrid approach that combines architectural priors, regularization, and replay policies, leading to state-of-the-art performance across various scenarios. Similarly, Kirichenko et al. [204] introduced a generative-discriminative hybrid model that leverages generative replay and functional regularization to avoid forgetting while detecting task changes, providing strong performance on CL benchmarks. Together, these hybrid methods offer a comprehensive solution to CL challenges by leveraging the strengths of both regularization and replay strategies [201–204].

In the energy domain, hybrid regularization and replay methods can be applied to improve the accuracy of load forecasting models under dynamic grid conditions. For instance, a model predicting energy demand could store a subset of historical data from prior seasons and replay it during training to reinforce its understanding of long-term consumption trends while learning new patterns, such as those caused by extreme weather events. Regularization techniques could further stabilize the learned relationships between weather variables and energy demand, preventing performance degradation when adapting to emerging scenarios like renewable energy integration or shifts in consumption behavior. This dual approach ensures robustness and adaptability, crucial for maintaining grid stability and optimizing energy distribution.

7.1.4. Dynamic task-specific learning rates

Investigating dynamic learning rates tailored to the importance of tasks or developing task-aware optimizers could help manage the trade-off between retaining stability and enhancing plasticity. Dynamic Task-Specific Learning Rates (DTSLR) improve CL by adjusting the learning rates based on the task's complexity and the model's progress on each task [205,206]. This approach allows models to dynamically adapt their learning speed, leading to better retention of previous tasks while still efficiently learning new ones. For instance, in models like Dynamic Sparse Distributed Memory (DSDM), which focuses on non-stationary data and task-free CL, DTSLR can help balance learning new tasks and retaining previously learned knowledge by controlling how much weight updates impact different network parts. This adjustment reduces catastrophic forgetting by allowing more gradual learning in areas related to older tasks while speeding up learning for new, unseen data, which improves memory efficiency and task-specific performance over time [207]. In other frameworks, such as DualNet, DTSLR can ensure optimal performance between fast and slow learning systems by modulating learning rates based on how fast or slow knowledge should be retained, enhancing the effectiveness of CL across various domains [208].

In the energy domain, dynamic learning rates tailored to task importance could enhance the performance of renewable energy forecasting models. For example, DTSLR could be employed in a wind energy forecasting system where the model adjusts its learning speed based on seasonal variations in wind patterns. Higher learning rates could accelerate adaptation to rapidly changing conditions during peak wind months, while lower rates during stable periods could help preserve long-term knowledge about average wind behavior. Similarly, DTSLR can optimize model updates in dynamic grid management systems by gradually learning the impact of renewable integration while retaining knowledge about baseline grid performance. This approach would reduce catastrophic forgetting, ensure reliable predictions, and improve overall energy system resilience in evolving conditions.

7.1.5. Constrained optimization frameworks

Future work could formalize the stability–plasticity dilemma as a constrained optimization problem, aiming to minimize loss on new tasks while constraining the loss on old tasks. Research could focus on improving such optimization methods to balance learning across tasks without drastic performance losses [209,210]. A constrained optimization approach could minimize forecasting errors for new energy sources (e.g., solar farms coming online) while maintaining accuracy for existing sources (e.g., wind farms). For instance, integrating these models into a hybrid energy grid must avoid disrupting operational stability when introducing new data. Physics-Informed Neural Networks (PINNs) also seem to be a promising solution that combines historical knowledge or known principles with data-driven models for more accurate and reliable estimation [211].

7.2. Overcoming data scarcity problems

7.2.1. Self-supervised and semi-supervised learning

Self-supervised and semi-supervised learning techniques are critical in improving CL by enabling models to learn from unlabeled data while mitigating catastrophic forgetting. Self-Supervised Learning (SSL) creates strong feature representations without requiring labeled datasets, particularly useful when data is introduced sequentially in CL scenarios.

Studies such as [212,213] demonstrated that by incorporating SSL objectives like Barlow twins or distillation-based methods into the CL framework, the models can retain previously learned knowledge while adapting to new tasks. These methods help maintain the quality of learned representations, avoid retraining from scratch, and optimize resource efficiency. Moreover, semi-supervised learning techniques allow for the combination of both labeled and unlabeled data, which is vital in real-world scenarios where fully labeled datasets are scarce, as highlighted in [214,215]. By using representation learning and designing loss functions that balance knowledge retention and new data acquisition, SSL-based CL models achieve comparable or superior performance to traditional supervised models, especially in applications like earth observation and sound recognition. This helps create scalable and flexible CL systems that generalize better across different domains, including the energy domain, where large volumes of unlabeled data are generated continuously from sensors, smart meters, and renewable energy systems. For example, in energy load forecasting, SSL techniques could create robust representations of seasonal and daily consumption patterns without requiring fully labeled historical data. Similarly, semi-supervised approaches could integrate labeled data from specific grid nodes with unlabeled data from new or under-monitored areas, improving model adaptability while preserving learned knowledge about prior consumption trends. These techniques are particularly valuable for optimizing renewable energy integration and managing dynamic energy grids, where data variability and scarcity of labels often challenge traditional supervised learning approaches.

7.2.2. Few-shot and zero-shot learning

Advancements in Few-Shot Learning (FSL) and Zero-Shot Learning (ZSL) have significant implications for improving CL, especially in scenarios where labeled data is limited or unavailable. FSL allows models to generalize from minimal examples, which is crucial in CL tasks where acquiring a large amount of labeled data for each new task is impractical. For instance, as noted in [216], incorporating prototype augmentation and multi-teacher knowledge transfer mechanisms in Continual ZSL (CZSL) helps strike a balance between maintaining stability on old tasks and maintaining the flexibility to learn new ones. FSL techniques can help models quickly adapt to new tasks with few examples, reducing negative transfer and enhancing the retention of past knowledge. Similarly, ZSL contributes to CL by enabling models to generalize to unseen tasks or categories without explicitly training on them, leveraging semantic information like class attributes or representations. For example, [217,218] demonstrated that generative models and meta-learning approaches can enhance learning new tasks without prior access to unseen classes. These advancements allow for faster training and more efficient adaptation in CZSL, reducing catastrophic forgetting while expanding the model's capacity to handle new tasks. By leveraging FSL and ZSL, CL models can generalize more effectively and efficiently in dynamic environments where the data distribution evolves.

7.2.3. Data augmentation techniques for CL

In the energy domain, leveraging image-based frameworks such as 2D-CNNs and PConv for transforming time-series data into images enables innovative approaches like image-inpainting models to handle missing energy data, improving data imputation accuracy and supporting downstream tasks such as energy forecasting and building energy modeling [219]. Extending data augmentation and synthetic data generation techniques like GANs and VAEs to CL tasks can provide more training data, particularly in domains with limited availability. Developing domain-specific augmentation methods may improve model robustness and generalization [220].

7.2.4. Task embedding for knowledge transfer

Task knowledge transfer and embedding for knowledge transfer are critical in improving CL, particularly in scenarios with limited data or highly specialized tasks. By enabling knowledge transfer across tasks, these methods can help models learn new tasks more effectively while mitigating catastrophic forgetting. The idea is to extract transferable task embeddings or representations from previously learned tasks and apply them to new tasks, allowing the model to generalize even with limited task-specific data. In the energy domain, this could transfer knowledge from models trained on specific energy systems, such as solar power generation, to related tasks like wind energy prediction or energy load forecasting, enabling more efficient learning with limited labeled data. For example, in [221], the authors proposed a novel CL model that leverages pre-trained models for knowledge transfer across tasks, enhancing task performance and reducing catastrophic forgetting. This approach ensures that valuable knowledge from previous tasks is retained and reused for future tasks, optimizing the model's capacity to handle sequential learning tasks. Similarly, [222] introduced a biologically inspired method for dynamically managing knowledge transfer, selectively forgetting old knowledge that hinders the learning of new tasks, which further improves forward knowledge transfer. Such dynamic methods could be applied to scenarios like energy data imputation, where older patterns in seasonal energy usage are selectively refined to accommodate newer data trends. Moreover, [223] highlighted how Adaptive Knowledge Transfer with a Multi-classifier Ensemble (AKTME) can improve fault diagnosis by effectively distilling shared knowledge across multiple auxiliary tasks. In the context of energy systems, this could translate to fault detection in smart grids by leveraging auxiliary tasks like anomaly detection in energy consumption patterns. This CL framework adapts pre-learned representations to new tasks, allowing the model to recognize rare conditions with limited data. This highlights how task embeddings can be continuously refined and transferred, making the model more resilient in data-scarce environments and improving task performance in CL systems.

7.3. Promoting security and privacy

7.3.1. Privacy-preserving CL

Developing privacy-preserving CL is essential for future research, particularly when sensitive data, such as medical records or user data, is used. As CL involves learning sequentially from different tasks, there is a significant risk of sensitive information leakage across tasks, especially when data from previous tasks is retained or accessed to prevent catastrophic forgetting. Implementing robust privacy-preserving mechanisms like differential privacy or FL can help mitigate these risks while maintaining model performance. For instance, in healthcare, as highlighted in [224], CL can mitigate the need for frequent retraining by using privacy-preserving methods to protect patient data. This study demonstrated that privacy-preserving CL algorithms can perform well even without retaining prior patient data, thus offering a solution to the challenges of long-term deployment of AI models in clinical settings. Similarly, in [225], the Dream Net model for face emotion recognition ensures privacy using a pseudo-rehearsal approach, which helps preserve privacy without storing explicit examples of previously learned data. Moreover, in decentralized learning environments like FL, privacy risks are amplified as data is distributed across multiple clients. Privacy-preserving methods such as federated clustering, as explored by [226], and differentially private FL, as introduced in [227], help protect client data by ensuring that sensitive information is not leaked across tasks or devices. These methods can adapt to changing privacy requirements and handle non-iid data distributions, enabling more secure and scalable CL solutions. Privacy-preserving CL in the energy domain can be applied through FL to optimize energy usage across smart grids while ensuring that individual household energy data remains confidential. Additionally, differentially private mechanisms can be used in renewable energy forecasting, enabling models to adapt to new weather patterns without compromising sensitive location-specific energy consumption data.

7.3.2. Adversarial robustness in CL

CL systems are prone to adversarial vulnerabilities, where models are manipulated by carefully crafted perturbations, undermining their reliability in dynamic environments. Adversarial attacks in the energy domain include injecting perturbations into smart meter data to manipulate energy usage predictions or tampering with renewable energy forecasting models to create inaccurate supply–demand estimates. Research into adversarial robustness in CL emphasizes the importance of robust feature extraction to enhance the security of models against attacks. For instance, [189] showed that utilizing robust features over non-robust ones can significantly improve adversarial resilience in class-IL, as robust features make the model less susceptible to noise and adversarial attacks, compared to non-robust or mixed-feature models. Another study by [228] introduced Task-Aware Boundary Augmentation (TABA), a novel method that strengthens CL models against adversarial attacks by dynamically adapting decision boundaries for different tasks, proving its efficacy in experiments on CIFAR-10 and CIFAR-100. However, there are deeper concerns regarding adversarial vulnerabilities, as illustrated by [229], who demonstrated that CL systems are particularly vulnerable to adversarial attacks across all task-incremental, domain-incremental, and class-incremental settings. They emphasize the ease with which adversaries can target any learned task, highlighting a critical challenge in securing continually learned knowledge. To address this, approaches such as Retrospective Adversarial Replay (RAR) proposed by [230] introduced adversarial perturbations into replay buffers, refining the boundary between old and new tasks, which mitigates catastrophic forgetting while bolstering the model's robustness against adversarial inputs. Additionally, the susceptibility of CL models to backdoor attacks, as highlighted by [231], revealed that even a small percentage of manipulated data can compromise the integrity of a continually learned model, injecting false memories or malicious misinformation. This poses significant threats in real-world applications where CL models must operate securely over time, making it essential for future research to explore adaptive adversarial defense strategies that evolve with the model's CL process. Robustness in CL in the energy domain can be achieved by using techniques like adversarial training, which incorporates adversarial examples during model updates, or by leveraging TABA to dynamically adapt decision boundaries for evolving tasks (e.g., energy demand forecasting, under varying weather conditions). Additionally, integrating mechanisms such as RAR can mitigate catastrophic forgetting while enhancing resilience to adversarial inputs in dynamic energy systems (e.g., in energy grids that include renewable resources frequently added or removed from the grid).

7.3.3. Federated CL

Federated CL (FCL) combines the strengths of FL and CL to address the challenge of learning from non-stationary data in decentralized, privacy-preserving systems such as energy management and IoT applications. By allowing distributed clients to learn sequential tasks from private data streams without sharing raw data, FCL preserves privacy while enabling ongoing learning. However, this approach presents challenges like task interference, communication overhead, and data heterogeneity, which must be addressed for effective implementation. For example, task interference may occur when an FCL model trained on renewable energy forecasting for solar farms struggles to adapt to wind energy forecasting tasks, as the features and patterns from solar data can conflict with those needed for accurate wind predictions. One solution to task interference and inefficient knowledge sharing in FCL is Federated Weighted Inter-client Transfer (FedWeIT), introduced by [232]. FedWeIT addresses these issues by separating global federated parameters from task-specific parameters, allowing clients to share only relevant knowledge. This minimizes the negative impact of unrelated tasks, enhances overall task performance, and significantly reduces communication costs—critical in decentralized systems. This selective knowledge transfer ensures that FCL can effectively adapt to new tasks while preventing catastrophic forgetting. In environments with constrained resources, such as edge devices in IoT applications, efficient and fast learning is essential. To tackle this, [233] proposed ADMM-FedMeta, a federated meta-learning framework designed for continual edge learning. By leveraging prior task knowledge and using an ADMM-based approach to reduce computational overhead, ADMM-FedMeta allows rapid adaptation to new tasks. This makes it well-suited for real-time decision-making in edge applications, where quick CL is required without overloading the system. Additionally, data heterogeneity, where data distributions differ across clients and over time, is a major challenge in FCL. As discussed by [234], CL strategies are particularly well-suited to handle this issue, as they are designed to cope with shifting data streams. Integrating these strategies into FL systems can improve performance by ensuring models adapt to evolving data. Finally, secure aggregation techniques are crucial in FCL for maintaining privacy while ensuring that model updates from multiple clients are aggregated without compromising individual data. This is particularly important in applications like energy management, where privacy and security are paramount but robust, and CL is still necessary to improve system performance over time.

7.3.4. Zero-knowledge proofs

ZKPs offer a cryptographic method to enhance security and privacy in CL by allowing one party (the prover) to prove the correctness of a statement (such as a model's learning outcome) without revealing sensitive information to the verifier. This is especially useful in decentralized environments like FL and ML as a Service (MLaaS), where data privacy is critical. In FL, ZKPs can safeguard against malicious behavior by central aggregators. For instance, [235] introduced zkFL, a framework that uses ZKPs to verify each aggregation round without revealing local model details, protecting against tampering. ZkFL integrates blockchain technology to validate the proofs securely, ensuring privacy and security without altering the FL architecture or affecting speed. Additionally, [236] proposed zk Proof of Training (zkPoT), which allows model owners to prove that they have trained their models correctly without exposing the training data. This is crucial in scenarios like outsourced ML, ensuring that models are trained as specified while maintaining data privacy. ZkPoT can efficiently verify models such as logistic regression, and its techniques could extend to more complex models. ZKPs are also vital in ensuring the integrity of model prediction services. In MLaaS environments, as discussed by [237], ZKPs can verify the accuracy of predictions without exposing model parameters or input data, ensuring that the service remains secure and trustworthy. ZKPs further enhance cryptographic primitives, such as lattice-based cryptography, as proposed by [238], improving performance while maintaining security. They also extend to applications like decentralized credential verification, as seen in CrossCert [239], ensuring secure, anonymous credential checks in distributed networks. ZKPs techniques can enhance CL in the energy domain by enabling secure and privacy-preserving verification of model updates or learning outcomes. For example, ZKPs could be used in decentralized energy management systems to prove that a CL model has accurately learned to optimize energy distribution based on private, evolving data streams from smart meters without exposing the underlying data. Additionally, ZKP frameworks like zkFL can ensure that model updates in FCL are aggregated securely, safeguarding against malicious tampering while maintaining the privacy of individual clients in tasks such as renewable energy forecasting or load balancing.

7.4. Improving scalability and interoperability

7.4.1. Efficient distributed learning frameworks

Future research should explore distributed and decentralized learning frameworks tailored to CL scenarios. Such systems could help scale learning across multiple devices or cloud-based systems, enabling large-scale task learning without overburdening individual models.

7.4.2. Modular and task-agnostic architectures

Developing modular architectures that compartmentalize learning for different tasks can promote scalability, efficiency, and interoperability in CL. Such architectures can allow the system to efficiently handle new tasks without redundant learning by reusing or freezing specific modules dedicated to prior tasks. This approach facilitates the isolation of task-specific knowledge, thereby enhancing the system's ability to mitigate catastrophic forgetting and enabling forward knowledge transfer. For instance, [240] proposed a Mixture-of-Variational-Experts layer that creates information processing paths through the network, governed by a gating policy. This modular approach ensures that each sub-network specializes in specific tasks, reducing forgetting by leveraging diverse and specialized parameters while maintaining the flexibility to learn new tasks in a task-agnostic manner. Additionally, methods like the one introduced in [241] utilized a LOss DEcoupling (LODE) strategy, which separates learning objectives for new tasks, providing a more balanced trade-off between stability (retaining knowledge) and plasticity (learning new tasks). Similarly, [242,243] emphasize task-agnostic architectures where models can incrementally learn tasks without requiring explicit task identities during inference. These approaches promote the reuse of shared task-agnostic features while adapting to specific task-related features. For example, the Task Agnostic Representation Consolidation (TARC) method combined task-agnostic self-supervised learning with task-specific supervised learning to maintain generalizable representations across tasks. This modular and compartmentalized learning approach allows models to dynamically adjust to new tasks without redundant retraining, significantly improving CL's flexibility and robustness. In the energy domain, such modular and task-agnostic architectures could be applied to smart grid management, where task-specific modules handle energy demand forecasting, anomaly detection, or renewable energy integration, while shared task-agnostic components ensure seamless adaptation to emerging tasks like EV charging optimization or real-time energy trading without disrupting existing functionalities.

7.4.3. Interoperability via standardized model protocols

Developing standardized frameworks for model interoperability in CL is essential for enabling efficient knowledge sharing and collaboration across different domains. These frameworks promote scalability, adaptability, and performance in decentralized or federated systems by allowing models to communicate and exchange learned knowledge without redundant retraining. For instance, tools like MMdnn, introduced in [244], provide a unified Intermediate Representation (IR) for converting models between different DL frameworks while preserving their semantic integrity. This approach can be extended to CL environments to ensure seamless platform interoperability. Moreover, solutions like the Lion-based Shuffled Shepherd (Lion-SS) optimization algorithm proposed in [245] for application migration in cloud environments offer valuable insights for optimizing resource allocation in CL systems. By leveraging such optimization techniques, CL models can be efficiently migrated or shared across various platforms, ensuring compatibility and performance.

7.4.4. Efficient task allocation

Efficient task allocation in CL improves scalability and model performance as new tasks are learned over time. Continual Task Allocation via Sparse Prompting (CoTASP), introduced by [246], enhances this process by utilizing sparse prompts to allocate sub-networks within a meta-policy network. This method allows related tasks to share resources efficiently, preventing catastrophic forgetting while ensuring optimal model capacity. By dynamically adjusting network weights and prompts, CoTASP achieves a balance between learning new tasks (plasticity) and retaining prior knowledge (stability). Task prioritization is another key aspect, as demonstrated by [247]. Their approach identifies cooperative relations between tasks, allowing models to focus on helpful information from past tasks, reducing interference, and improving learning efficiency. Additionally, in Mobile Edge Computing (MEC) environments, where resources are limited, efficient task allocation is crucial. Zhan et al. [248] proposed a CL-based resource allocation method that optimized system performance by adapting to user demand without requiring full retraining, reducing computational and energy costs. In the energy domain, efficient task allocation techniques could be applied to dynamically manage energy resources in smart grids, where tasks like renewable energy integration, demand forecasting, and fault detection are prioritized and allocated based on their interdependencies, ensuring optimal performance and reducing the need for costly retraining or redundant computations.

7.5. Improving practical applications of CL in energy systems

Advancing CL in energy systems requires focused efforts in several key areas. Developing scalable architectures, such as modular, task-specific, and dynamic designs, can facilitate scalability and task-sharing across diverse energy infrastructures [120,124]. Hybrid learning models that integrate supervised, semi-supervised, and self-supervised techniques are essential for addressing the challenge of limited labeled data in real-world settings [103,115]. Enhancing security and privacy through FCL, adversarial robustness, and privacy-preserving methods like ZKPs is critical for protecting sensitive energy data [119,137]. Leveraging domain-specific knowledge, such as energy conservation laws and physical principles, can improve model robustness and ensure practical applicability [106,163]. Lastly, rigorous benchmarking and real-world case studies are necessary to validate CL methods and address implementation gaps, demonstrating their feasibility across diverse energy scenarios [30,106].

In this regard, future research in energy systems should prioritize hybrid CL approaches that integrate replay methods and architectural solutions to handle fluctuating renewable energy outputs, incorporate climate-specific data for improved adaptability, and enhance scalability for larger distributed systems [124,125]. In building energy management, privacy-preserving CL techniques such as FL, modular architectures for building-specific adaptation, and lifelong learning systems for continuous adaptation to evolving conditions are essential [103,105]. Additionally, in energy trading and market optimization, future efforts should focus on developing adversarially robust CL models, employing ZKP methods for secure data sharing, and integrating blockchain technology to ensure tamper-proof transactions and enhance system reliability [124,137].

8. Conclusion

In summary, this paper thoroughly and comprehensively examined different methods of CL, reviewed various studies that utilized CL in the energy sector, and conducted a case study on detecting anomalies in energy usage. After highlighting the main challenges that could hinder the deployment of CL, the paper suggested potential directions for improved deployment of CL. To expand, we thoroughly carried out the following activities:

- Examined the literature, focusing on approaches like replay, regularization tactics, and architectural methods that tackle the issues related to catastrophic forgetting.
- Surveyed various energy-related CL applications, including non-intrusive load monitoring, demand-side management, fault/anomaly detection, load forecasting/prediction, and renewable energy integration.
- Conducted a case study about identifying anomalies in energy systems and compared various CL methods.
- Identified key challenges that could hinder the deployment of CL in the energy domain and specified prospective approaches to address those challenges.

According to the case study's conclusions, replay-based strategies performed best on both new and old tasks in terms of average accuracy and forgetting rate. The best course of action would be to combine the regularization-based method with earlier samples replayed during training. Future work will concentrate on employing hybrid CL approaches and observing their performance. Additionally, we plan to conduct the anomaly detection case study in a semi-supervised or even unsupervised manner, providing valuable insights into the robustness of the CL methodologies.

CRediT authorship contribution statement

Aya Nabil Sayed: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Yassine Himeur:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Conceptualization. **Iraklis Varlamis:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Conceptualization. **Faycal Bensaali:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper was made possible by National Priorities Research Program (NPRP) grant No. NPRP14S-0401-210122 from the Qatar Research Development and Innovation Council (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors. Open Access funding provided by the Qatar National Library.

Data availability

The data used is publicly available.

References

[1] Liu J, Zhou Y, Yang H, Wu H. Uncertainty energy planning of net-zero energy communities with peer-to-peer energy trading and green vehicle storage considering climate changes by 2050 with machine learning methods. Appl Energy 2022;321:119394.

[2] Elnour M, Fadli F, Himeur Y, Petri I, Rezgui Y, Meskin N, Ahmad AM. Performance and energy optimization of building automation and management systems: Towards smart sustainable carbon-neutral sports facilities. Renew Sustain Energy Rev 2022;162:112401.

[3] Hamilton I, Kennard H, Rapf O, Amorocho J, Steuwer S, Kockat J, Toth Z, et al. Global Status Report for Buildings and Construction-Beyond foundations: Mainstreaming sustainable solutions to cut emissions from the buildings sector. United Nations Environment Programme; 2024.

[4] Change IC, et al. Mitigation of climate change. In: Contribution of working group III to the fifth assessment report of the intergovernmental panel on climate change. Vol. 1454, 2014, p. 147.

[5] Olu-Ajayi R, Alaka H, Sulaimon I, Sunmola F, Ajayi S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. J Build Eng 2022;45:103406.

[6] Himeur Y, Ghanem K, Alsalemi A, Bensaali F, Amira A. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. Appl Energy 2021;287:116601.

[7] Shin W, Han J, Rhee W. AI-assistance for predictive maintenance of renewable energy systems. Energy 2021;221:119775.

[8] Elnour M, Himeur Y, Fadli F, Mohammedsherif H, Meskin N, Ahmad AM, Petri I, Rezgui Y, Hodorog A. Neural network-based model predictive control system for optimizing building automation and management systems of sports facilities. Appl Energy 2022;318:119153.

[9] Shekhar H, Bhushan Mahato C, Suman SK, Singh S, Bhagyalakshmi L, Prasad Sharma M, Laxmi Kantha B, Agraharam SK, Rajaram A. Demand side control for energy saving in renewable energy resources using deep learning optimization. Electr Power Compon Syst 2023;51(19):2397–413.

[10] Himeur Y, Elnour M, Fadli F, Meskin N, Petri I, Rezgui Y, Bensaali F, Amira A. AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. Artif Intell Rev 2023;56(6):4929–5021.

[11] Himeur Y, Rimal B, Tiwary A, Amira A. Using artificial intelligence and data fusion for environmental monitoring: A review and future perspectives. Inf Fusion 2022;86:44–75.

[12] Copiaco A, Himeur Y, Amira A, Mansoor W, Fadli F, Atalla S, Sohail SS. An innovative deep anomaly detection of building energy consumption using energy time-series images. Eng Appl Artif Intell 2023;119:105775.

[13] De Lange M, Aljundi R, Masana M, Parisot S, Jia X, Leonardis A, Slabaugh G, Tuytelaars T. A continual learning survey: Defying forgetting in classification tasks. IEEE Trans Pattern Anal Mach Intell 2022;44(7):3366–85. http://dx.doi.org/10.1109/TPAMI.2021.3057446.

[14] Parisi GI, Kemker R, Part JL, Kanan C, Wermter S. Continual lifelong learning with neural networks: A review. Neural Netw 2019;113:54–71. http://dx.doi.org/10.1016/j.neunet.2019.01.012.

[15] van de Ven G, Tuytelaars T, Tolias A. Three types of incremental learning. Nat Mach Intell 2022;4:1–13. http://dx.doi.org/10.1038/s42256-022-00568-3.

[16] Aljundi R, Rohrbach M, Tuytelaars T. Selfless sequential learning. 2019, arXiv:1806.05421.

[17] Parisi GI, Kemker R, Part JL, Kanan C, Wermter S. Continual lifelong learning with neural networks: A review. Neural Netw 2019;113:54–71.

[18] Hadsell R, Rao D, Rusu AA, Pascanu R. Embracing change: Continual learning in deep neural networks. Trends in Cognitive Sciences 2020;24(12):1028–40.

[19] De Lange M, Aljundi R, Masana M, Parisot S, Jia X, Leonardis A, Slabaugh G, Tuytelaars T. A continual learning survey: Defying forgetting in classification tasks. IEEE Trans Pattern Anal Mach Intell 2021;44(7):3366–85.

[20] Van de Ven GM, Tuytelaars T, Tolias AS. Three types of incremental learning. Nat Mach Intell 2022;4(12):1185–97.

[21] Wickramasinghe B, Saha G, Roy K. Continual learning: A review of techniques, challenges and future directions. IEEE Trans Artif Intell 2023.

[22] Wang L, Zhang X, Su H, Zhu J. A comprehensive survey of continual learning: theory, method and application. IEEE Trans Pattern Anal Mach Intell 2024.

[23] Lesort T, Lomonaco V, Stoian A, Maltoni D, Filliat D, Díaz-Rodríguez N. Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. Inf Fusion 2020;58:52–68.

[24] Shaheen K, Hanif MA, Hasan O, Shafique M. Continual learning for real-world autonomous systems: Algorithms, challenges and frameworks. J Intell Robot Syst 2022;105(1):9.

[25] Mai Z, Li R, Jeong J, Quispe D, Kim H, Sanner S. Online continual learning in image classification: An empirical survey. Neurocomputing 2022;469:28–51.

[26] Hurtado J, Salvati D, Semola R, Bosio M, Lomonaco V. Continual learning for predictive maintenance: Overview and challenges. Intell Syst Appl 2023;19:200251.

[27] Kudithipudi D, Daram A, Zyarah A, Zohora FT, Aimone J, Yanguas-Gil A, Soures N, Neftci E, Mattina M, Lomonaco V, Thiem C, Epstein B. Design principles for lifelong learning AI accelerators. Nat Electron 2023;6. http://dx.doi.org/10.1038/s41928-023-01054-3.

[28] Graffieti G, Borghi G, Maltoni D. Continual learning in real-life applications. IEEE Robot Autom Lett 2022;7(3):6195–202.

[29] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Int J Surg 2010;8(5):336–41. http://dx.doi.org/10.1016/j.ijsu.2010.02.007.

[30] Li A, Zhang C, Xiao F, Fan C, Deng Y, Wang D. Large-scale comparison and demonstration of continual learning for adaptive data-driven building energy prediction. Appl Energy 2023;347:121481.

[31] Lomonaco V, Maltoni D. Core50: a new dataset and benchmark for continuous object recognition. In: Conference on robot learning. PMLR; 2017, p. 17–26.

[32] Nguyen CV, Achille A, Lam M, Hassner T, Mahadevan V, Soatto S. Toward understanding catastrophic forgetting in continual learning. 2019, arXiv preprint arXiv:1908.01091.

[33] Zhao X, Wang H, Huang W, Lin W. A statistical theory of regularization-based continual learning. 2024, arXiv preprint arXiv:2406.06213.

[34] Zhang Y, Lin Z, Sun Y, Yin F, Fritsche C. Regularization-based efficient continual learning in deep state-space models. 2024, arXiv preprint arXiv:2403.10123.

[35] Nokhwal S, Kumar N. Rtra: Rapid training of regularization-based approaches in continual learning. In: 2023 10th international conference on soft computing & machine intelligence. ISCMI, IEEE; 2023, p. 188–92.

[36] Li H, Wu J, Braverman V. Fixed design analysis of regularization-based continual learning. In: Conference on lifelong learning agents. PMLR; 2023, p. 513–33.

[37] Yang X, Li F, Li T, Ji W, Liang Y. Elastic weight consolidation continual learning based signal detection in multiple channel MIMO system. In: 2021 IEEE/CIC international conference on communications in China. ICCC, IEEE; 2021, p. 534–9.

[38] Li Z, Hoiem D. Learning without forgetting. IEEE Trans Pattern Anal Mach Intell 2017;40(12):2935–47.
[39] Li X, Wang S, Sun J, Xu Z. Variational data-free knowledge distillation for continual learning. IEEE Trans Pattern Anal Mach Intell 2023;45(10):12618–34.
[40] Rolnick D, Ahuja A, Schwarz J, Lillicrap T, Wayne G. Experience replay for continual learning. Adv Neural Inf Process Syst 2019;32.
[41] Atkinson C, McCane B, Szymanski L, Robins A. Pseudo-rehearsal: Achieving deep reinforcement learning without catastrophic forgetting. Neurocomputing 2021;428:291–307.
[42] Ji Z, Liu J, Wang Q, Zhang Z. Coordinating experience replay: A harmonious experience retention approach for continual learning. Knowl-Based Syst 2021;234:107589. http://dx.doi.org/10.1016/j.knosys.2021.107589.
[43] Khan V, Cygert S, Twardowski B, Trzciński T. Looking through the past: better knowledge retention for generative replay in continual learning. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 3496–500.
[44] Lopez-Paz D, Ranzato M. Gradient episodic memory for continual learning. Adv Neural Inf Process Syst 2017;30.
[45] Chaudhry A, Ranzato M, Rohrbach M, Elhoseiny M. Efficient lifelong learning with a-gem. 2018, arXiv preprint arXiv:1812.00420.
[46] Xu J, Ma J, Gao X, Zhu Z. Adaptive progressive continual learning. IEEE Trans Pattern Anal Mach Intell 2021;44(10):6715–28.
[47] Mallya A, Lazebnik S. Packnet: Adding multiple tasks to a single network by iterative pruning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 7765–73.
[48] Serra J, Suris D, Miron M, Karatzoglou A. Overcoming catastrophic forgetting with hard attention to the task. In: International conference on machine learning. PMLR; 2018, p. 4548–57.
[49] Lin S, Yang L, Fan D, Zhang J. Beyond not-forgetting: Continual learning with backward knowledge transfer. Adv Neural Inf Process Syst 2022;35:16165–77.
[50] Chen H-J, Cheng A-C, Juan D-C, Wei W, Sun M. Mitigating forgetting in online continual learning via instance-aware parameterization. Adv Neural Inf Process Syst 2020;33:17466–77.
[51] Verwimp E, Yang K, Parisot S, Hong L, McDonagh S, Pérez-Pellitero E, De Lange M, Tuytelaars T. Clad: A realistic continual learning benchmark for autonomous driving. Neural Netw 2023;161:659–69.
[52] Ghunaim Y, Bibi A, Alhamoud K, Alfarra M, Al Kader Hammoud HA, Prabhu A, Torr PH, Ghanem B. Real-time evaluation in online continual learning: A new hope. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 11888–97.
[53] Razdaibiedina A, Mao Y, Hou R, Khabsa M, Lewis M, Almahairi A. Progressive prompts: Continual learning for language models. 2023, arXiv preprint arXiv:2301.12314.
[54] Smith JS, Tian J, Halbe S, Hsu Y-C, Kira Z. A closer look at rehearsal-free continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 2410–20.
[55] McDonnell MD, Gong D, Parvaneh A, Abbasnejad E, van den Hengel A. Ranpac: Random projections and pre-trained models for continual learning. Adv Neural Inf Process Syst 2024;36.
[56] Gao Q, Zhao C, Sun Y, Xi T, Zhang G, Ghanem B, Zhang J. A unified continual learning framework with general parameter-efficient tuning. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 11483–93.
[57] Villa A, Alcázar JL, Alfarra M, Alhamoud K, Hurtado J, Heilbron FC, Soto A, Ghanem B. Pivot: Prompting for video continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 24214–23.
[58] Smith JS, Karlinsky L, Gutta V, Cascante-Bonilla P, Kim D, Arbelle A, Panda R, Feris R, Kira Z. Coda-prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 11909–19.
[59] Wang L, Xie J, Zhang X, Huang M, Su H, Zhu J. Hierarchical decomposition of prompt-based continual learning: Rethinking obscured sub-optimality. Adv Neural Inf Process Syst 2024;36.
[60] Lin H, Zhang B, Feng S, Li X, Ye Y. Pcr: Proxy-based contrastive replay for online class-incremental continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 24246–55.
[61] Kim S, Noci L, Orvieto A, Hofmann T. Achieving a better stability-plasticity trade-off via auxiliary networks in continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 11930–9.
[62] Heng A, Soh H. Selective amnesia: A continual learning approach to forgetting in deep generative models. Adv Neural Inf Process Syst 2024;36.
[63] Gao R, Liu W. Ddgr: Continual learning with deep diffusion-based generative replay. In: International conference on machine learning. PMLR; 2023, p. 10744–63.
[64] Zheng Z, Ma M, Wang K, Qin Z, Yue X, You Y. Preventing zero-shot transfer degradation in continual learning of vision-language models. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 19125–36.
[65] Wang L, Zhang X, Li Q, Zhang M, Su H, Zhu J, Zhong Y. Incorporating neuro-inspired adaptability for continual learning in artificial intelligence. Nat Mach Intell 2023;5(12):1356–68.
[66] Dohare S, Hernandez-Garcia JF, Rahman P, Mahmood AR, Sutton RS. Maintaining plasticity in deep continual learning. 2023, arXiv preprint arXiv:2306.13812.
[67] Wei Y, Ye J, Huang Z, Zhang J, Shan H. Online prototype learning for online continual learning. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 18764–74.
[68] Khan MGZA, Naeem MF, Van Gool L, Stricker D, Tombari F, Afzal MZ. Introducing language guidance in prompt-based continual learning. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 11463–73.
[69] Zhang G, Wang L, Kang G, Chen L, Wei Y. Slca: Slow learner with classifier alignment for continual learning on a pre-trained model. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 19148–58.
[70] Davalas C, Michail D, Diou C, Varlamis I, Tserpes K. Computationally efficient rehearsal for online continual learning. In: International conference on image analysis and processing. Springer; 2022, p. 39–49.
[71] Tanoni G, Principi E, Mandoliniy L, Squartini S. Appliance-incremental learning for non-intrusive load monitoring. In: 2023 IEEE international conference on communications, control, and computing technologies for smart grids. smartgridcomm, IEEE; 2023, p. 1–6.
[72] Sykiotis S, Kaselimi M, Doulamis A, Doulamis N. Continilm: A continual learning scheme for non-intrusive load monitoring. In: ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing. ICASSP, IEEE; 2023, p. 1–5.
[73] Li S, Zhao R, Li K, Lu J, Ren Y, Yu Z. Incremental learning for appliance identification based on VI trajectory signature. In: 2023 8th international conference on power and renewable energy. ICPRE, IEEE; 2023, p. 1467–71.
[74] Li Y, Li K, Wang S, Cao Z. Incremental learning based on VI trajectory feature for non-intrusive load monitoring. In: 2023 IEEE 7th conference on energy internet and energy system integration. EI2, IEEE; 2023, p. 5341–6.
[75] Zhang J, Tan Z, Li K, Hu B, Zhang J, Zhang R. Incremental classifier learning for non-intrusive appliance identification. In: 2024 9th Asia conference on power and electrical engineering. ACPEE, IEEE; 2024, p. 1298–302.
[76] Qiu L, Yu T, Lan C. A semi-supervised load identification method with class incremental learning. Eng Appl Artif Intell 2024;131:107768.
[77] Palensky P, Dietrich D. Demand side management: Demand response, intelligent energy systems, and smart loads. IEEE Trans Ind Inf 2011;7(3):381–8.
[78] Kim S, Mowakeaa R, Kim S-J, Lim H. Building energy management for demand response using kernel lifelong learning. IEEE Access 2020;8:82131–41.
[79] Kim S-J, Mowakeaa R. Kernel-based efficient lifelong learning algorithm. In: 2019 IEEE data science workshop. DSW, IEEE; 2019, p. 175–9.
[80] Hossain MB, Pokhrel SR, Choi J. Orchestrating smart grid demand response operations with URLLC and MuZero learning. IEEE Internet Things J 2023.
[81] Taboga V, Bellahsen A, Dagdougui H. Deep reinforcement learning for peak load reduction in aggregated residential houses. In: 2020 IEEE power & energy society general meeting. PESGM, IEEE; 2020, p. 1–5.
[82] Lee C, Kim S-H, Youn C-H. An accelerated continual learning with demand prediction based scheduling in edge-cloud computing. In: 2020 international conference on data mining workshops. ICDMW, IEEE; 2020, p. 717–22.
[83] Wu H, Elizaveta D, Zhadan A, Petrosian O. Forecasting online adaptation methods for energy domain. Eng Appl Artif Intell 2023;123:106499.
[84] Tang C, Qin Y, Wu F, Tang Z. Dynamic demand-aware power grid intelligent pricing algorithm based on deep reinforcement learning. IEEE Access 2024.
[85] Nguyen TD, Hua C-H, Rehman UU, Hussain M, Lee S. CLEB: A continual learning energy bidding framework for an energy market bidding application. In: Proceedings of the 2023 6th artificial intelligence and cloud computing conference. 2023, p. 46–53.
[86] Ng WW, Zhang J, Lai CS, Pedrycz W, Lai LL, Wang X. Cost-sensitive weighting and imbalance-reversed bagging for streaming imbalanced and concept drifting in electricity pricing classification. IEEE Trans Ind Inf 2018;15(3):1588–97.
[87] Hu B, Hao Z, Chen Z, Zhang J. Combination with continual learning update scheme for power system transient stability assessment. Sensors 2022;22(22):8982.
[88] Li X, Yang Z, Guo P, Cheng J. An intelligent transient stability assessment framework with continual learning ability. IEEE Trans Ind Inf 2021;17(12):8131–41.
[89] Liu J, Peng X, Wu L, Li Y, Liu J, Niu S, Huo C, Wei P, Ren C. Incremental learning method for online transient stability assessment of power system based on KKT condition. In: 2021 IEEE sustainable power and energy conference. iSPEC, IEEE; 2021, p. 3968–74.
[90] Ren C, Xu Y, Zhao J, Zhang R, Wan T. A super-resolution perception-based incremental learning approach for power system voltage stability assessment with incomplete PMU measurements. CSEE J Power Energy Syst 2021;8(1):76–85.
[91] Tian Y, Wang K, Oluic M, Ghandhari M, Xu J, Li G. Construction of multi-state transient stability boundary based on broad learning. IEEE Trans Power Syst 2020;36(4):2906–17.
[92] Li N, Li B, Han Y, Gao L. Dual cost-sensitivity factors-based power system transient stability assessment. IET Gener Transm Distrib 2020;14(24):5858–69.
[93] Cui J, Qin W, Zhu Z, Lu R, Qin P, Li X. A digital twin framework for transient stability assessment based on knowledge distillation. In: 2024 9th Asia conference on power and electrical engineering. ACPEE, IEEE; 2024, p. 2535–9.

[94] Amutha A, Uthra RA, Roselyn JP, Brunet RG. Anomaly detection and classification in streaming PMU data in smart grids. Comput Syst Sci Eng 2023;46(3):3387–401.

[95] Sifat AI, McFadden FJS, Rayudu R, Bailey J. Classification of stages of a high impedance fault using sequential learning algorithms. In: 2020 IEEE kansas power and energy conference. KPEC, IEEE; 2020, p. 1–6.

[96] Luo L, Wang J, Zhou S, Lou G, Sun J. A broad learning-based state estimation method for power system. Energy Rep 2022;8:1227–35.

[97] Alves Hd, Bretas AS, Bretas NG. Smart grids cyber-attack defense: A solution based on an incremental learning support vector machine. In: North American power symposium. 2019, p. 99–114.

[98] Mollaiee A, Ameli MT, Azad S. Novel continuous learning scheme for online static security assessment based on the weather-dependent security index. IET Gener Transm Distrib 2022;16(18):3684–705.

[99] Veerakumar N, Cremer JL, Popov M. Dynamic Incremental Learning for real-time disturbance event classification. Int J Electr Power Energy Syst 2023;148:108988.

[100] Zhu C, Pu Y, Lyu Z, Yang K, Yang Q. Multidimensional information fusion and broad learning system-based condition recognition for energy pipeline safety. Knowl-Based Syst 2024;300:112259.

[101] Cai G, Gao H, Yang D, Wang L. Synchrophasor-based online load margin estimation using incremental learning assisted LightGBM in smart grid. IEEE Syst J 2023;17(3):4660–70.

[102] Zhou Y, Tian X, Zhang C, Zhao Y, Li T. Elastic weight consolidation-based adaptive neural networks for dynamic building energy load prediction modeling. Energy Build 2022;265:112098.

[103] Prabowo A, Chen K, Xue H, Sethuvenkatraman S, Salim FD. Navigating Out-of-Distribution Electricity Load Forecasting during COVID-19: Benchmarking energy load forecasting models without and with continual learning. In: Proceedings of the 10th ACM international conference on systems for energy-efficient buildings, cities, and transportation. 2023, p. 41–50.

[104] Fekri MN, Patel H, Grolinger K, Sharma V. Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network. Appl Energy 2021;282:116177.

[105] Hu X, Zhang X, Zeng Y, Feng C, Meng D. Dominant parameter identification for virtual load model based on incremental learning. In: 2020 IEEE 4th conference on energy internet and energy system integration. EI2, IEEE; 2020, p. 3728–32.

[106] He Y, Henze J, Sick B. Continuous learning of deep neural networks to improve forecasts for regional energy markets. IFAC-Pap 2020;53(2):12175–82.

[107] Aragón G, Puri H, Grass A, Chala S, Beecks C. Incremental deep-learning for continuous load prediction in energy management systems. In: 2019 IEEE milan powertech. IEEE; 2019, p. 1–6.

[108] Jagait RK, Fekri MN, Grolinger K, Mir S. Load forecasting under concept drift: Online ensemble learning with recurrent neural network and ARIMA. IEEE Access 2021;9:98992–9008.

[109] Yang E, Lee C, Kim J-H, Tao TM, Youn C-H. StreamDL: Deep learning serving platform for AMI stream forecasting. In: 2020 international conference on data mining workshops. ICDMW, IEEE; 2020, p. 723–8.

[110] Besanger Y, Tran QT, et al. Self-updating machine learning system for building load forecasting-method, implementation and case-study on COVID-19 impact. Sustain Energy Grids Netw 2022;32:100873.

[111] Ramos D, Faria P, Vale Z, Correia R. Electricity consumption forecast in an industry facility to support production planning update in short time. In: 2020 IEEE international conference on environment and electrical engineering and 2020 IEEE industrial and commercial power systems europe. EEEIC/I&CPS europe, IEEE; 2020, p. 1–6.

[112] Ramos D, Faria P, Vale Z, Mourinho J, Correia R. Industrial facility electricity consumption forecast using artificial neural networks and incremental learning. Energies 2020;13(18):4774.

[113] Zamee MA, Han D, Won D. Online hour-ahead load forecasting using appropriate time-delay neural network based on multiple correlation–multicollinearity analysis in IoT energy network. IEEE Internet Things J 2021;9(14):12041–55.

[114] Kim S-H, Lee C, Youn C-H. An accelerated edge cloud system for energy data stream processing based on adaptive incremental deep learning scheme. IEEE Access 2020;8:195341–58.

[115] Zhang L, Zhu J, Zhang D, Liu Y. An incremental photovoltaic power prediction method considering concept drift and privacy protection. Appl Energy 2023;351:121919.

[116] Chupong C, Junhuathon N, Dangeam S, Plangklang B. Comparison of deep learning and incremental learning model for net load forecasting. ECTI Trans Electr Eng Electron Commun 2024;22(1).

[117] Chen S, Ge W, Liang X, Jin X, Du Z. Lifelong learning with deep conditional generative replay for dynamic and adaptive modeling towards net zero emissions target in building energy system. Appl Energy 2024;353:122189.

[118] da Silva MA, Abreu T, Santos-Júnior CR, Minussi CR. Load forecasting for smart grid based on continuous-learning neural network. Electr Power Syst Res 2021;201:107545.

[119] Yanmei J, Mingsheng L, Yangyang L, Yaping L, Jingyun Z, Yifeng L, Chunyang L. Enhanced neighborhood node graph neural networks for load forecasting in smart grid. Int J Mach Learn Cybern 2024;15(1):129–48.

[120] Ng RW, Begam KM, Rajkumar RK, Wong YW, Chong LW. An improved self-organizing incremental neural network model for short-term time-series load prediction. Appl Energy 2021;292:116912.

[121] Chupong C, Plangklang B. Incremental learning model for load forecasting without training sample. Comput Mater Contin 2022;72(3).

[122] Chupong C, Junhuathon N, Plangklang B. Short-term load forecasting by FOS-ELM with re-learning method. In: 2022 international conference on power, energy and innovations. ICPEI, IEEE; 2022, p. 1–4.

[123] Goh SM, Kow KW, Tan M, Rajkumar R, Wong YW. Hardware implementation of an active learning self-organizing neural network to predict the power fluctuation events of a photovoltaic grid-tied system. Microprocess Microsyst 2022;90:104448.

[124] Peng X, Wang B, Zhang L, Su P. Adaptive online sequential extreme learning machine with kernels for online ship power prediction. Energies 2021;14(17):5371.

[125] Bisoi R, Dash DR, Dash P, Tripathy L. An efficient robust optimized functional link broad learning system for solar irradiance prediction. Appl Energy 2022;319:119277.

[126] Puah BK, Chong LW, Wong YW, Begam K, Khan N, Juman MA, Rajkumar RK. A regression unsupervised incremental learning algorithm for solar irradiance prediction. Renew Energy 2021;164:908–25.

[127] Almaraash M, Abdulrahim M, Hagras H. A life-long learning XAI metaheuristic-based type-2 fuzzy system for solar radiation modelling. IEEE Trans Fuzzy Syst 2023.

[128] Sarmas E, Strompolas S, Marinakis V, Santori F, Bucarelli MA, Doukas H. An incremental learning framework for photovoltaic production and load forecasting in energy microgrids. Electronics 2022;11(23):3962.

[129] Hu T, Wu W, Guo Q, Sun H, Shi L, Shen X. Very short-term spatial and temporal wind power forecasting: A deep learning approach. CSEE J Power Energy Syst 2019;6(2):434–43.

[130] He Y, Sick B. CLeaR: An adaptive continual learning framework for regression tasks. AI Perspect 2021;3(1):2.

[131] Khan N, Haq IU, Ullah FUM, Khan SU, Lee MY. CL-net: ConvLSTM-based hybrid architecture for batteries' state of health and power consumption forecasting. Mathematics 2021;9(24):3326.

[132] Fu Y, Howard A, Zeng C, Chen Y, Gao P, Stinis P. Physics-guided continual learning for predicting emerging aqueous organic redox flow battery material performance. ACS Energy Lett 2024;9:2767–74.

[133] You D, Gao S, Chen X. Incremental learning-based online risk assessment of power system operation with high renewable energy penetration. In: 2023 IEEE sustainable power and energy conference. iSPEC, IEEE; 2023, p. 1–6.

[134] Yang Y-D, Yao W-J, Yang L-Z. Evaluation and adjustment based on broad learning system for small signal stability analysis of power system integrated with wind farms. J Electr Eng Technol 2022;17(6):3107–18.

[135] Khademhamedani B, Izadi M, Izadi N, Azarshab A, Najari A, Zaker B. Transformative control optimization in PMSG-based wind energy systems: A deep reinforcement learning approach. In: 2023 3rd international conference on electrical machines and drives. ICEMD, IEEE; 2023, p. 1–6.

[136] Zhang R, Xu Y. Data-driven dynamic security assessment and control of power systems: an online sequential learning method. J Energy Eng 2019;145(5):04019019.

[137] Li M, Mour N, Smith L. Machine learning based on reinforcement learning for smart grids: Predictive analytics in renewable energy management. Sustain Cities Soc 2024;109:105510.

[138] Himeur Y, Alsalemi A, Bensaali F, Amira A. A novel approach for detecting anomalous energy consumption based on micro-moments and deep neural networks. Cogn Comput 2020;12. http://dx.doi.org/10.1007/s12559-020-09764-y.

[139] Uttama Nambi AS, Reyes Lua A, Prasad VR. Loced: Location-aware energy disaggregation framework. In: Proceedings of the 2nd acm international conference on embedded systems for energy-efficient built environments. 2015, p. 45–54.

[140] Himeur Y, Alsalemi A, Bensaali F, Amira A. Building power consumption datasets: Survey, taxonomy and future directions. Energy Build 2020;227:110404.

[141] Sayed AN, Bensaali F, Himeur Y, Houchati M. Image transformation approaches for occupancy detection: A comprehensive analysis. In: The proceedings of the international conference on smart city applications. Springer; 2023, p. 303–11.

[142] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 770–8.

[143] Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, et al. Overcoming catastrophic forgetting in neural networks. Proc Natl Acad Sci 2017;114(13):3521–6.

[144] Zenke F, Poole B, Ganguli S. Continual learning through synaptic intelligence. In: International conference on machine learning. PMLR; 2017, p. 3987–95.

[145] Aljundi R, Babiloni F, Elhoseiny M, Rohrbach M, Tuytelaars T. Memory aware synapses: Learning what (not) to forget. In: Proceedings of the European conference on computer vision. ECCV, 2018, p. 139–54.

[146] Lomonaco V, Maltoni D, Pellegrini L, et al. Rehearsal-free continual learning over small non-IID batches. In: CVPR workshops. Vol. 1, 2020, p. 3.

[147] Lomonaco V, Pellegrini L, Cossu A, Carta A, Graffieti G, Hayes TL, De Lange M, Masana M, Pomponi J, Van de Ven GM, et al. Avalanche: an end-to-end library for continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021, p. 3600–10.
[148] Carta A, Pellegrini L, Cossu A, Hemati H, Lomonaco V. Avalanche: A pytorch library for deep continual learning. J Mach Learn Res 2023;24(363):1–6.
[149] Ali B, Javed M, et al. Data incremental learning for fault detection and classification in photovoltaic arrays. In: 2021 IEEE 13th international symposium on diagnostics for electrical machines, power electronics and drives. SDEMPED, Vol. 1, IEEE; 2021, p. 310–5.
[150] Gurbuz MB, Dovrolis C. Nispa: Neuro-inspired stability-plasticity adaptation for continual learning in sparse networks. 2022, arXiv preprint arXiv:2206.09117.
[151] Liang Y-S, Li W-J. Adaptive plasticity improvement for continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 7816–25.
[152] Mishra P, Narayanan R. Stable continual learning through structured multiscale plasticity manifolds. Curr Opin Neurobiol 2021;70:51–63.
[153] Lee A, Gomes HM, Zhang Y. Balancing the stability-plasticity dilemma with online stability tuning for continual learning. In: 2022 international joint conference on neural networks. IJCNN, IEEE; 2022, p. 1–8.
[154] Li J, Yu T, Zhang X, Li F, Lin D, Zhu H. Efficient experience replay based deep deterministic policy gradient for AGC dispatch in integrated energy system. Appl Energy 2021;285:116386.
[155] De Lange M, van de Ven G, Tuytelaars T. Continual evaluation for lifelong learning: Identifying the stability gap. 2022, arXiv preprint arXiv:2205.13452.
[156] Kim D, Han B. On the stability-plasticity dilemma of class-incremental learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 20196–204.
[157] Dohare S, Hernandez-Garcia JF, Lan Q, Rahman P, Mahmood AR, Sutton RS. Loss of plasticity in deep continual learning. Nature 2024;632(8026):768–74.
[158] Chen Q, Shui C, Han L, Marchand M. On the stability-plasticity dilemma in continual meta-learning: theory and algorithm. Adv Neural Inf Process Syst 2024;36.
[159] Hammami Z, Sayed-Mouchaweh M, Mouelhi W, Ben Said L. Neural networks for online learning of non-stationary data streams: a review and application for smart grids flexibility improvement. Artif Intell Rev 2020;53(8):6111–54.
[160] Thangarasa V, Miconi T, Taylor GW. Enabling continual learning with differentiable hebbian plasticity. In: 2020 international joint conference on neural networks. IJCNN, IEEE; 2020, p. 1–8.
[161] Prabowo A, Chen K, Xue H, Sethuvenkatraman S, Salim FD. Continually learning out-of-distribution spatiotemporal data for robust energy forecasting. In: Joint European conference on machine learning and knowledge discovery in databases. Springer; 2023, p. 3–19.
[162] Schug S, Benzing F, Steger A. Presynaptic stochasticity improves energy efficiency and helps alleviate the stability-plasticity dilemma. Elife 2021;10:e69884.
[163] Yang Y, Huang J, Hu D. Lifelong learning with shared and private latent representations learned through synaptic intelligence. Neural Netw 2023;163:165–77.
[164] Tercan H, Deibert P, Meisen T. Continual learning of neural networks for quality prediction in production using memory aware synapses and weight transfer. J Intell Manuf 2022;33(1):283–92.
[165] Wang Q, Wang R, Li Y, Wei D, Wang H, Ma K, Zheng Y, Meng D. Relational experience replay: Continual learning by adaptively tuning task-wise relationship. IEEE Trans Multimed 2024.
[166] Thandiackal K, Portenier T, Giovannini A, Gabrani M, Goksel O. Generative feature-driven image replay for continual learning. Image Vis Comput 2024;150:105187.
[167] Pomponi J, Scardapane S, Lomonaco V, Uncini A. Efficient continual learning in neural networks with embedding regularization. Neurocomputing 2020;397:139–48.
[168] Chrysakis A, Moens M-F. Online continual learning from imbalanced data. In: International conference on machine learning. PMLR; 2020, p. 1952–61.
[169] Lin S, Ju P, Liang Y, Shroff N. Theory on forgetting and generalization of continual learning. In: International conference on machine learning. PMLR; 2023, p. 21078–100.
[170] Gao Y, Ruan Y, Fang C, Yin S. Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data. Energy Build 2020;223:110156.
[171] Jiang Y, Liu Y, Zhan W. Overcoming catastrophic forgetting with detail-degradation decoupling networks for super-resolution. In: Journal of physics: conference series, vol. 2637, IOP Publishing; 2023, 012023.
[172] Wang R, Rayhana R, Gholami M, Herrera OE, Liu Z, Mérida W. Multi-task deep learning for large-scale buildings energy management. Energy Build 2024;307:113964.
[173] Niu R, Wu Z, Song C. Representation space maintenance: Against forgetting in continual learning. In: 2024 international joint conference on neural networks. IJCNN, IEEE; 2024, p. 1–7.
[174] Xu J, Li K, Li D. Multioutput framework for time-series forecasting in smart grid meets data scarcity. IEEE Trans Ind Inf 2024.
[175] Lin R, Yu C, Liu T. Eliminating catastrophic overfitting via abnormal adversarial examples regularization. Adv Neural Inf Process Syst 2024;36.
[176] Caccia L, Pineau J. Special: Self-supervised pretraining for continual learning. In: International workshop on continual semi-supervised learning. Springer; 2021, p. 91–103.
[177] Fan C, Lei Y, Sun Y, Mo L. Novel transformer-based self-supervised learning methods for improved HVAC fault diagnosis performance with limited labeled data. Energy 2023;278:127972.
[178] Zhang Y, Zhou G, Liu Z, Huang L, Ren Y. An adaptive transfer learning framework for data-scarce HVAC power consumption forecasting. IEEE Trans Sustain Energy 2024.
[179] Lin W, Wu D, Jenkin M. Electric load forecasting for individual households via spatial-temporal knowledge distillation. IEEE Trans Power Syst 2024.
[180] Xu J, Li K, Li D. An automated few-shot learning for time series forecasting in smart grid under data scarcity. IEEE Trans Artif Intell 2024.
[181] Veniat T, Denoyer L, Ranzato M. Efficient continual learning with modular networks and task-driven priors. 2020, arXiv preprint arXiv:2012.12631.
[182] Ruan J, Wang Q, Chen S, Lyu H, Liang G, Zhao J, Dong ZY. On vulnerability of renewable energy forecasting: Adversarial learning attacks. IEEE Trans Ind Inf 2023.
[183] Zhao AP, Zhang Q, Alhazmi M, Hu PJ-H, Zhang S, Yan X. AI for science: Covert cyberattacks on energy storage systems. J Energy Storage 2024;99:112835.
[184] Chen W, Zhou A, Zhou P, Gao L, Ji S, Wu D. A privacy-preserving online learning approach for incentive-based demand response in smart grid. IEEE Syst J 2019;13(4):4208–18.
[185] Badr MM, Ibrahem MI, Mahmoud M, Alasmary W, Fouda MM, Almotairi KH, Fadlullah ZM. Privacy-preserving federated-learning-based net-energy forecasting. In: SoutheastCon 2022. IEEE; 2022, p. 133–9.
[186] Agah N, Mohammadi J, Aved A, Ferris D, Cruz EA, Morrone P. Data poisoning: An overlooked threat to power grid resilience. 2024, arXiv preprint arXiv:2407.14684.
[187] Zhao J, Jung T, Wang Y, Li X. Achieving differential privacy of data disclosure in the smart grid. In: IEEE INFOCOM 2014-IEEE conference on computer communications. IEEE; 2014, p. 504–12.
[188] Alam N, Rahman M, Islam MR, Hossain M. Ensemble adversarial training-based robust model for multi-horizon dynamic line rating forecasting against adversarial attacks. Electr Power Syst Res 2025;241:111289.
[189] Khan H, Bouaynaya NC, Rasool G. Adversarially robust continual learning. In: 2022 international joint conference on neural networks. IJCNN, IEEE; 2022, p. 1–8.
[190] Xie M, Ren K, Lu Y, Yang G, Xu Q, Wu B, Lin J, Ao H, Xu W, Shu J. Kraken: memory-efficient continual learning for large-scale real-time recommendations. In: SC20: international conference for high performance computing, networking, storage and analysis. IEEE; 2020, p. 1–17.
[191] de Castilho RE, Ide N, Kim J-D, Klie J-C, Suderman K. Towards cross-platform interoperability for machine-assisted text annotation. Genom Inform 2019;17(2).
[192] Ntafalias A, Tsakanikas S, Skarvelis-Kazakos S, Papadopoulos P, Skarmeta-Gómez AF, González-Vidal A, Tomat V, Ramallo-González AP, Marin-Perez R, Vlachou MC. Design and implementation of an interoperable architecture for integrating building legacy systems into scalable energy management systems. Smart Cities 2022;5(4):1421–40.
[193] Yang Z, Zhang S, Li C, Wang M, Wang H, Zhang M. Efficient knowledge management for heterogeneous federated continual learning on resource-constrained edge devices. Future Gener Comput Syst 2024;156:16–29.
[194] Manca F, Ratto F, Palumbo F. ONNX-to-hardware design flow for adaptive neural-network inference on FPGAs. 2024, arXiv preprint arXiv:2406.09078.
[195] Gupta G, Yadav K, Paull L. Look-ahead meta learning for continual learning. Adv Neural Inf Process Syst 2020;33:11588–98.
[196] Javed K, White M. Meta-learning representations for continual learning. Adv Neural Inf Process Syst 2019;32.
[197] Han Y-n, Liu J-w. Online continual learning via the meta-learning update with multi-scale knowledge distillation and data augmentation. Eng Appl Artif Intell 2022;113:104966.
[198] Son J, Lee S, Kim G. When meta-learning meets online and continual learning: A survey. IEEE Trans Pattern Anal Mach Intell 2024.
[199] Faber K, Corizzo R, Sniezynski B, Japkowicz N. Lifelong continual learning for anomaly detection: New challenges, perspectives, and insights. IEEE Access 2024;12:41364–80.
[200] Liu B, Mazumder S. Lifelong and continual learning dialogue systems: learning during conversation. In: Proceedings of the AAAI conference on artificial intelligence. Vol. 35, 2021, p. 15058–63.
[201] Merlin G, Lomonaco V, Cossu A, Carta A, Bacciu D. Practical recommendations for replay-based continual learning methods. In: International conference on image analysis and processing. Springer; 2022, p. 548–59.
[202] Han X, Guo Y. Continual learning with dual regularizations. In: Machine learning and knowledge discovery in databases. Research track: European conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, proceedings, part i 21. Springer; 2021, p. 619–34.

[203] Lomonaco V, Pellegrini L, Graffieti G, Maltoni D. Architect, regularize and replay: A flexible hybrid approach for continual learning. In: Towards human brain inspired lifelong learning. World Scientific; 2024, p. 101–21.

[204] Kirichenko P, Farajtabar M, Rao D, Lakshminarayanan B, Levine N, Li A, Hu H, Wilson AG, Pascanu R. Task-agnostic continual learning with hybrid probabilistic models. 2021, arXiv preprint arXiv:2106.12772.

[205] Madotto A, Lin Z, Zhou Z, Moon S, Crook P, Liu B, Yu Z, Cho E, Wang Z. Continual learning in task-oriented dialogue systems. 2020, arXiv preprint arXiv:2012.15504.

[206] Yildirim MO, Gok EC, Sokar G, Mocanu DC, Vanschoren J. Continual learning with dynamic sparse training: Exploring algorithms for effective model updates. In: Conference on parsimony and learning. PMLR; 2024, p. 94–107.

[207] Pourcel J, Vu N-S, French RM. Online task-free continual learning with dynamic sparse distributed memory. In: European conference on computer vision. Springer; 2022, p. 739–56.

[208] Pham Q, Liu C, Hoi S. Dualnet: Continual learning, fast and slow. Adv Neural Inf Process Syst 2021;34:16131–44.

[209] Mundt M, Hong Y, Pliushch I, Ramesh V. A wholistic view of continual learning with deep neural networks: Forgotten lessons and the bridge to active and open world learning. Neural Netw 2023;160:306–36.

[210] Fajemisin AO, Maragno D, den Hertog D. Optimization with constraint learning: A framework and survey. European J Oper Res 2024;314(1):1–14.

[211] Ngo Q-H, Nguyen BL, Vu TV, Zhang J, Ngo T. Physics-informed graphical neural network for power system state estimation. Appl Energy 2024;358:122602.

[212] Fini E, Da Costa VGT, Alameda-Pineda X, Ricci E, Alahari K, Mairal J. Self-supervised models are continual learners. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022, p. 9621–30.

[213] Marsocci V, Scardapane S. Continual barlow twins: continual self-supervised learning for remote sensing semantic segmentation. IEEE J Sel Top Appl Earth Obs Remote Sens 2023;16:5049–60.

[214] Wang Z, Subakan C, Jiang X, Wu J, Tzinis E, Ravanelli M, Smaragdis P. Learning representations for new sound classes with continual self-supervised learning. IEEE Signal Process Lett 2022;29:2607–11.

[215] Tang CI, Qendro L, Spathis D, Kawsar F, Mascolo C, Mathur A. Kaizen: Practical self-supervised continual learning with continual fine-tuning. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2024, p. 2841–50.

[216] Lu J, Sun S. PAMK: Prototype augmented multi-teacher knowledge transfer network for continual zero-shot learning. IEEE Trans Image Process 2024.

[217] Verma V, Mehta N, Liang KJ, Mishra A, Carin L. Meta-learned attribute self-interaction network for continual and generalized zero-shot learning. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2024, p. 2721–31.

[218] Zhang W, Janson P, Yi K, Skorokhodov I, Elhoseiny M. Continual zero-shot learning through semantically guided generative random walks. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 11574–85.

[219] Fu C, Quintana M, Nagy Z, Miller C. Filling time-series gaps using image techniques: Multidimensional context autoencoder approach for building energy data imputation. Appl Therm Eng 2024;236:121545.

[220] Du K, Lyu F, Li L, Hu F, Feng W, Xu F, Xi X, Cheng H. Multi-label continual learning using augmented graph convolutional network. IEEE Trans Multimed 2023.

[221] Ke Z, Liu B, Ma N, Xu H, Shu L. Achieving forgetting prevention and knowledge transfer in continual learning. Adv Neural Inf Process Syst 2021;34:22443–56.

[222] Wang L, Zhang M, Jia Z, Li Q, Bao C, Ma K, Zhu J, Zhong Y. Afec: Active forgetting of negative transfer in continual learning. Adv Neural Inf Process Syst 2021;34:22379–91.

[223] Xing S, Lei Y, Yang B, Lu N. Adaptive knowledge transfer by continual weighted updating of filter kernels for few-shot fault diagnosis of machines. IEEE Trans Ind Electron 2021;69(2):1968–76.

[224] Verma T, Jin L, Zhou J, Huang J, Tan M, Choong BCM, Tan TF, Gao F, Xu X, Ting DS, et al. Privacy-preserving continual learning methods for medical image classification: a comparative analysis. Front Med 2023;10:1227515.

[225] Mainsant M, Solinas M, Reyboz M, Godin C, Mermillod M. Dream net: a privacy preserving continual leaming model for face emotion recognition. In: 2021 9th international conference on affective computing and intelligent interaction workshops and demos. ACIIW, IEEE; 2021, p. 01–8.

[226] Masuyama N, Nojima Y, Toda Y, Loo CK, Ishibuchi H, Kubota N. Privacy-preserving continual federated clustering via adaptive resonance theory. IEEE Access 2024.

[227] Chathoth AK, Necciai CP, Jagannatha A, Lee S. Differentially private federated continual learning with heterogeneous cohort privacy. In: 2022 IEEE international conference on big data. big data, IEEE; 2022, p. 5682–91.

[228] Bai T, Chen C, Lyu L, Zhao J, Wen B. Towards adversarially robust continual learning. In: ICASSP 2023-2023 IEEE international conference on acoustics, speech and signal processing. ICASSP, IEEE; 2023, p. 1–5.

[229] Khan H, Shah PM, Zaidi SFA, Zia Q, et al. Susceptibility of continual learning against adversarial attacks. 2022, arXiv preprint arXiv:2207.05225.

[230] Kumari L, Wang S, Zhou T, Bilmes JA. Retrospective adversarial replay for continual learning. Adv Neural Inf Process Syst 2022;35:28530–44.

[231] Umer M, Polikar R. Adversarial targeted forgetting in regularization and generative based continual learning models. In: 2021 international joint conference on neural networks. IJCNN, IEEE; 2021, p. 1–8.

[232] Yoon J, Jeong W, Lee G, Yang E, Hwang SJ. Federated continual learning with weighted inter-client transfer. In: International conference on machine learning. PMLR; 2021, p. 12073–86.

[233] Yue S, Ren J, Xin J, Lin S, Zhang J. Inexact-ADMM based federated meta-learning for fast and continual edge learning. In: Proceedings of the twenty-second international symposium on theory, algorithmic foundations, and protocol design for mobile networks and mobile computing. 2021, p. 91–100.

[234] Criado MF, Casado FE, Iglesias R, Regueiro CV, Barro S. Non-iid data and continual learning processes in federated learning: A long road ahead. Inf Fusion 2022;88:263–80.

[235] Wang Z, Dong N, Sun J, Knottenbelt W, Guo Y. Zkfl: Zero-knowledge proof-based gradient aggregation for federated learning. IEEE Trans Big Data 2024.

[236] Garg S, Goel A, Jha S, Mahloujifar S, Mahmoody M, Policharla G-V, Wang M. Experimenting with zero-knowledge proofs of training. In: Proceedings of the 2023 ACM SIGSAC conference on computer and communications security. 2023, p. 1880–94.

[237] Fan Y, Xu B, Zhang L, Song J, Zomaya A, Li K-C. Validating the integrity of convolutional neural network predictions based on zero-knowledge proof. Inform Sci 2023;625:125–40.

[238] Lyubashevsky V, Nguyen NK, Plançon M. Lattice-based zero-knowledge proofs and applications: shorter, simpler, and more general. In: Annual international cryptology conference. Springer; 2022, p. 71–101.

[239] Tran T-D, Minh PK, Thuy TLT, Duy PT, Cam NT, Pham V-H. CrossCert: A privacy-preserving cross-chain system for educational credential verification using zero-knowledge proof. In: International conference on industrial networks and intelligent systems. Springer; 2024, p. 256–71.

[240] Hihn H, Braun DA. Hierarchically structured task-agnostic continual learning. Mach Learn 2023;112(2):655–86.

[241] Liang Y-S, Li W-J. Loss decoupling for task-agnostic continual learning. Adv Neural Inf Process Syst 2024;36.

[242] Bhat PS, Zonooz B, Arani E. Task agnostic representation consolidation: a self-supervised based continual learning approach. In: Conference on lifelong learning agents. PMLR; 2022, p. 390–405.

[243] Singh PR, Gopalakrishnan S, ZhongZheng Q, Suganthan PN, Ramasamy S, Ambikapathi A. Task-agnostic continual learning using base-child classifiers. In: 2021 IEEE international conference on image processing. ICIP, IEEE; 2021, p. 794–8.

[244] Liu Y, Chen C, Zhang R, Qin T, Ji X, Lin H, Yang M. Enhancing the interoperability between deep learning frameworks by model conversion. In: Proceedings of the 28th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering. 2020, p. 1320–30.

[245] Hiremath TC, KS R. Optimization enabled deep learning method in container-based architecture of hybrid cloud for portability and interoperability-based application migration. J Exp Theor Artif Intell 2024;36(7):985–1002.

[246] Yang Y, Zhou T, Jiang J, Long G, Shi Y. Continual task allocation in meta-policy network via sparse prompting. In: International conference on machine learning. PMLR; 2023, p. 39623–38.

[247] Jin H, Kim E. Helpful or harmful: Inter-task association in continual learning. In: European conference on computer vision. Springer; 2022, p. 519–35.

[248] Zhan S, Mao J, Chu Z, Xiong K, Fan P, Letaief KB. Multiresource allocation in dynamic MEC systems: a continual learning-based approach. In: Third international conference on electronic information engineering, big data, and computer technology. EIBDCT 2024, Vol. 13181, SPIE; 2024, p. 1931–40.