


Continual Learning for Time Series Forecasting: A First Survey [†]

Quentin Besnard * and Nicolas Ragot * Université de Tours, Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT),
37200 Tours, France

* Correspondence: quentin.besnard@univ-tours.fr (Q.B.); nicolas.ragot@univ-tours.fr (N.R.)

[†] Presented at the 10th International Conference on Time Series and Forecasting, Gran Canaria, Spain,
15–17 July 2024.

Abstract: Deep learning has brought significant advancements in the field of artificial intelligence, particularly in robotics, imaging, sound processing, etc. However, a common major challenge faced by all neural networks is their substantial demand for data during the learning process. The required data must be both quantitative and stationary to ensure the proper computing of standard models. Nevertheless, complying to these constraints is often impossible for many real-life applications because of dynamic environments. Indeed, modifications can occur in the distribution of the data or even in the goals to pursue within these environments. This is known as data and concept drift. Research in the field of continual learning seeks to address these challenges by implementing evolving models capable of adaptation over time. This notably involves finding a compromise on the plasticity/stability dilemma while taking into account material and computational constraints. Exploratory efforts are evident in all applications of deep learning (graphs, reinforcement learning, etc.), but to date, there is still a limited amount of work in the case of time series, specifically in the context of regression and forecasting. This paper aims to provide a first survey on this field of continuous learning applied to time series forecasting.

Keywords: continual learning; time series forecasting; deep learning; nonstationary environment; lifelong learning



Citation: Besnard, Q.; Ragot, N. Continual Learning for Time Series Forecasting: A First Survey. *Eng. Proc.* **2024**, *68*, 49. <https://doi.org/10.3390/engproc2024068049>

Academic Editors: Olga Valenzuela, Fernando Rojas, Luis Javier Herrera, Hector Pomares and Ignacio Rojas

Published: 17 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Machine learning applied to nonstationary environments, where changes can occur over time, is a rapidly growing field as it represents a challenge for future artificial intelligence applications. The exploration of new models with adaptive capabilities is currently grouped under the terminology of continual learning (CL). This domain encompasses themes like progressive learning [1], incremental learning [2], and lifelong learning [3]. Despite some specific differences, these various designations converge on the topic of a continuously learned model and the issue of catastrophic forgetting. Indeed, in real-world applications, obtaining an exhaustive and representative dataset before learning is often impossible. Thus, continual learning advantage lies in its incremental learning ability over time, avoiding complete retraining on a potentially substantial amount of data. Its interest, compared to standard models, is particularly significant in the case of data drifts where the data distribution changes over time [4]. Overall, the idea is to seek the perfect model able to evolve with nonstationary data, while dealing with plasticity/stability dilemma and without exhibiting symptoms of forgetting knowledge [5–10]. This is achieved while enhancing its efficiency when applied to new data as well as on previously seen data.

Several solutions exist for addressing the issue of catastrophic forgetting [11]. They can be grouped into three kind of strategies: rehearsal strategy, consisting of repeating a fraction of past data during learning; regularization-based strategy, applied on model parameters; and structural strategy, where the model grows over time when encountering

new tasks [12–16]. Each strategy and its hybrid/cross versions [1,17–19] present a trade-off between constraints and advantages.

The study of continual learning takes multiple research directions depending on application domains, including strategies dedicated to graphs [20], reinforcement learning [21,22], classification [23], or regression [7] tasks. In this work, we focus on the review of forecasting tasks applied on time series data, drawing a parallel with the classification domain, which represents one of the most explored axes in continual learning. Additionally, this work seeks to provide concise insights about required notions in the topic of continual learning while presenting compelling analyses aimed at addressing inherent questions associated with the principles of continual learning.

2. Continual Learning Principles

The problem of catastrophic forgetting, which is the cornerstone of continual learning, is correlated with the notion of plasticity/stability dilemma. A model exhibits great stability when it learns from a fixed dataset, supposed to represent perfectly and extensively all the situations (perfect sampling). If the data do not change over time, meaning their distribution, patterns, or the environment remain constant, then the stability of a model enables it to predict the desired outcome with high precision throughout all its usage. Nevertheless, it fails to address changes in data distribution or in goals to achieve. On the other side, the plasticity of a model allows it to continuously and quickly adapt to new tasks and data, which is one of the main strengths of continual learning. However, the downside of excessive plasticity is the loss of retaining past knowledge: as the model evolves over time, the accuracy decreases on previously seen data. This is the principle of catastrophic forgetting. For a standard model, a fine-tuning mechanism could be applied occasionally to allow model readjustment to some extent, for new tasks and data. Nevertheless, this strategy remains limited by the amount of data needed for adaptation as well as by the fact that old knowledge and abilities of the network are still lost [24]. The goal of continual learning is to address this plasticity/stability dilemma, but the solution is often constrained by an increasing cost in performance (memory, time, model size, etc.).

To address the catastrophic forgetting issue, the literature of continual learning introduces three strategies: data rehearsal, parameter regularization, and model extension [24]. The principle of data replay (rehearsal) involves constantly reintroducing samples of past data into learning steps to adapt the model [25]. The goal is to prevent forgetting by continuously stimulating the model with past knowledge. This approach often yields the best results in terms of model accuracy, but it comes with a perpetual increase in the amount of data that the model must ingest over time. Thus, recent research is focusing on a rehearsal-free version [26]. Multiple axes can be studied regarding the management of data to be reintroduced, as presented in [13,27–29]. The approach of architectural modification (model extension), presented in [30,31], involves adapting the model directly according to the needs. The size of the layers is growing while encountering new tasks, avoiding modification of parts of the network that are used for managing previous tasks. By avoiding the modification of previously learned weights, there is no loss in performance on previously learned tasks. However, similar to the data replay strategy, its downside is the continuous growth in the model size. This issue can be addressed by incorporating a pruning mechanism [32] to remove components within the network as they become unnecessary. But this process is not well established and still has its drawbacks regarding performance. Lastly, the regularization approach holds the best potential in the long run, although it may potentially exhibit some forgetting over time. It involves penalizing the modification of neural network weights based on their impact on model accuracy. As presented in the EWC algorithm [12,33,34], the aim is to select, through penalization, a weight space of the model that is the best to address all the tasks learned beforehand. The major advantage is that the work takes place in a Fisher matrix to compute the model loss during its update, which does not imply network or data replay growth. On the contrary, if too many disjointed tasks are learned, it can lead the model to propose an average version

of itself, satisfying a minimum performance on all tasks without performing in any of them. These three main strategies also have their hybrid versions, combining, for example, regularization with data replay to seek the best possible performances [17,19,35–37].

Currently, the field of continual learning research is mainly focused on classification tasks (imagery, natural language processing, medicine, etc.) and, particularly, on three scenarios for adapting the model over time, as presented in [2]:

- “Task Incremental Learning where it sequentially learns to solve a number of distinct tasks” (a task is associated with achieving a goal within a context, and if either or both change, we switch tasks);
- “Class Incremental Learning discriminate between incrementally observed classes”;
- “Domain Incremental Learning learn to solve the same problem in different contexts”.

In all these situations, the model is progressive, and its classification capabilities increase overall as it encounters new data over time. Thus, the goal within the classification research field is to propose new approaches to solve the catastrophic forgetting problem while achieving satisfactory precision results. Thanks to several standardized databases, such as MNIST and CIFAR, etc., used to compare and evaluate the performances of new models, new continual learning classification models [12,15,17,32] can be compared easily with already verified standard models.

3. Continual Learning for Time Series

Similarly to in the classification research field, continual learning applied to time series forecasting aims to address drifts in tasks and data, with the ability to continuously learn and make a model evolve over time, while resolving catastrophic forgetting issues. Although there are relatively common databases representing time series, most research on continuous learning for regression tasks focuses on specific applications for each use case which. Even if these applications promise concrete improvements in the field of time series forecasting, the comparison between standard models as well as other continuous learning algorithms is not straightforward.

3.1. Main Principles and Overview

Compared to classification domain, two scenarios can be encountered in the forecasting domain, as outlined in [38]—page 4:

- Incremental learning of the data domain, which refers to the situation where the underlying data generation process is changing over time, due to the nonstationarity of the data stream. This means that the distribution of the data relative to the same objective is varying over time.
- Incremental learning of the target domain, which refers to the situation where the output of the model varies over time. This is the case when the number or properties of prediction targets is changing (prediction of new variables like in multioutputs networks, changing in time prediction horizon, etc).

The literature on continual learning for regression and forecasting does not address, at all, the task incremental scenario, as defined previously for classification problems. When ‘task’ is mentioned, it most often refers to one of the two scenarios defined here (target or data domain incremental learning).

The incremental learning of the data domain refers either to tasks change defined by external supervision, human, or machine, as in the case of labels for classification, or to tasks that are drifting “naturally” from one to another because of changes in the environment that are not always identifiable (apparent drift vs. real drift). The fact is that most research works on continual learning for time series are considering task management based on an arbitrary separation of the dataset into fixed-size subsets, each one corresponding to a new task [24,27,29,39–46]. These approaches are called Task-based approaches. Although operational and useful for evaluation of CL systems, this task management may not represent real transitions between tasks. It mainly relies on an

incremental learning approach where the addition of tasks allows the model to progress through the data, leaving catastrophic forgetting management solely to the implemented strategy. On the contrary, correct task management (either based on explicit supervision of the system by an expert or based on an unsupervised supervision of the data) can help the model to avoid catastrophic forgetting (including better selection of data for replay in the rehearsal approach or proper selection of weights to retain in the regularization approach), as well as to deal more precisely with plasticity/stability dilemma. Ref. [47] particularly addresses the issues related to arbitrary task definition and proposes a new task detection system through loss analysis during model learning (referred to as Task-free approach). Other research works, such as [35], propose the idea of “novelty” data buffer based on their distance from the prediction accuracy.

Considering the application domain of continual learning for forecasting, we can observe that it is mainly applied in the energy domain [24,35,36,39–41,46]. Other sub-cases of application refer to industrial maintenance [36,43], climatic [36,42,45], or traffic analysis [27,36]. Each application uses a distinct dataset, hindering direct model comparison. It is noteworthy that, in all presented application cases, even the lowest performances still surpass those of standard baseline models. Also, we can observe that even if the issue of catastrophic forgetting [11] was already known, it was not widely emphasized for improving model performance. Only the last few years have seen a trend of consolidating research in this field under a common banner like continual learning, with the question of solving the ‘catastrophic forgetting’ matter.

3.2. Continual Learning Analysis

Continual learning techniques have already proven to be effective in various domains. In the case of forecasting data from time series, [24] highlights the interest of the principle (named here lifelong learning) (see Figure 1). It uses a mechanism of replaying sample from past data coming from the time series.

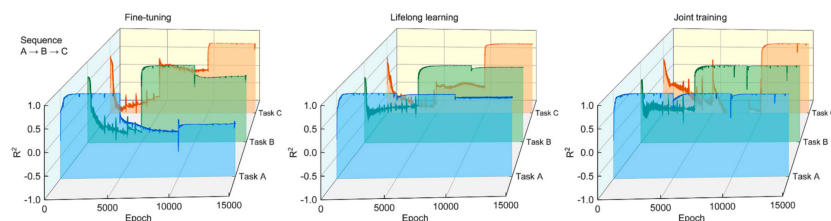


Figure 1. Comparison of fine-tuning, lifelong learning, and joint training for sequential task learning A (blue) → B (green) → C (red)—Figure 11— [24].

This approach is compared with fine-tuning and Joint-training, which provide an upper bound of performances in terms of prediction accuracy. The upper bound is represented by the joint dataset of all available learning tasks at a given time. Lifelong learning demonstrates performances close to the upper bound, contrary to fine-tuning, which exhibits a loss of precision on old tasks.

He et al. [41] (see Figure 2) also illustrate the superiority in term of MSE of some continual learning models (LWF, EWC, O-EWC, SI) compared to a baseline in the context of target and data-domain incremental scenarios. This is achieved most often at the expense of a cumulative increase in learning time. The dots show MSE metric and the learning time for each experiment. Squares indicate the average test error for each task during training, and stars indicate the average test error for each task after training all tasks.

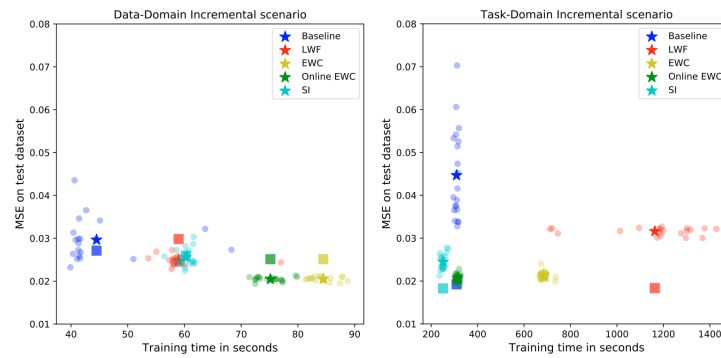


Figure 2. Comparison of training time and average MSE on test datasets over 20 experiments for algorithms in the data and target domain scenario—Figures 1 and 2— [41].

In the same paper [41] (see Figure 3), the authors provide an analysis of the forgetting rate (Equation (1)) along with the new tasks encountered in time series forecasting. The results show a better overall accuracy (Figure 2), as well as a much lower forgetting rate over time, when using continual learning models, compared to the standard model without any help against catastrophic forgetting (Figure 3).

$$forgetting\ ratio = \frac{\max(0, L_{warm_up^2} - L_{warm_up^1})}{L_{warm_up^1}}, \tag{1}$$

where $L_{warm_up^1}$ indicates the MSE on the warm-up dataset at the end of the warm-up phase and $L_{warm_up^2}$ indicates the error on the same dataset at the end of the update phase.

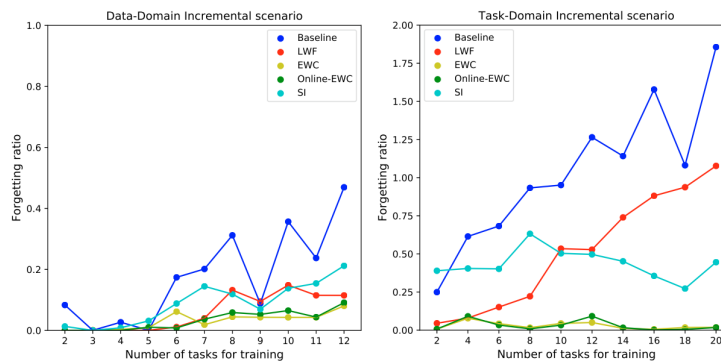


Figure 3. Performance of algorithms with increasing number of tasks by 2 in each step in the task and data domain incremental scenario—Figures 3 and 4— [41].

In the case of highly periodic datasets, the question of the usefulness of a continual model against a standard model arises. The idea is that a training on a sufficiently large dataset encompassing an entire periodicity may suffice on its own. Presented in [35] is an experiment based on an artificial dataset with daily and annual periodicity as well as an application on a dataset representing wind farm energy production (Experiment 1—Table 9 —[35]). The models were trained on a subset of data in an offline manner before being applied to online learning. On average, the results on periodic data showed worse performances than the standard model due to the plasticity of continual learning. Specifically, 1120 models were trained, and only 32.1% showed better results than the standard model. The parameterization of the CLear model [35] applied here defines its ability to understand the notion of data periodicity. With the right selection, the continual learning model can still be close to a standard model. That said, the application of the model to the second dataset from wind farms distinctly demonstrates the advantage that continual learning can bring to time series forecasting. Although the wind-dependent dataset exhibits daily and seasonal periodicity, climatic variations have applied variations within the data

shift. This can be observed in the results given in Table 1, where three instances, "A", "B", and "C", of the same model are compared: instance "C" represents the continual learning model, which adapts more effectively than the standard model "A" and the fine-tuned model "B" to changes. Furthermore, this study shows a lower average catastrophic forgetting rate, for instance, "C", both in the AE (autoencoder) part and in the predictor, then "B" (fine-tuning instance), highlighting the contribution of strategies to provide assistance against knowledge forgetting when learning (see Table 2).

Table 1. Average MSE for the three instance of the same model over 10 wind farms forecasting datasets (Table 9—[35]). Model C is the CLear continual learning model (best in bold).

Fitting Error MSE (e-2)				
	Instance A	Instance B	Instance C	Baseline
Mean	5.138	5.442	2.829	3.190

Table 2. Average forgetting ratio over the 10 wind farms forecasting datasets from CLear model instances (Table 11 —[35]).

Forgetting Ratio				
	Instance B (AE)	Instance C (AE)	Instance B	Instance C
Mean	1.402	1.171	3.550	1.161

3.3. Trends and Challenges

The development of deep learning in time series forecasting is still growing, with little concrete application cases [48]. However, emerging sub-domains like continual learning are already being studied for use in dynamic environments. The literature shows promising application outcomes designed to face the challenge of unstable environments where data evolve over time. Two main trends in the research can be seen:

- The first one focuses on model evolution and the crossover of different common continual learning strategies to more effectively mitigate catastrophic forgetting [35].
- The second distinguishes itself on a structural level by implementing models with a biologically inspired approach, aiming for behavioral mimicry of living organisms [36].

Future challenges revolve around adapting complex and recent popular deep models, such as those used in image or language processing. Notably, encoder–decoder systems and attention mechanisms [49] show great promise, with transformers or autoencoders that have proven their ability to handle time series data [24,35,36,41,43]. Furthermore, while it is important to diversify the application of such models and their strategies to different use cases to provide a broad range of expertise, a lack of common benchmarks in the literature on continual learning for time series prediction is noticeable and highlights the need for resolution of this.

Also, most strategies against catastrophic forgetting were initially developed for supervised classification tasks and then adapted for regression tasks. This explains the low number of research papers proposing task-free solutions, where tasks are not predefined (contrary to what is mainly seen in image classification). Also, papers that implement incremental learning models, like [50–55], do not adopt an approach against catastrophic forgetting and focus only on adaptive approaches over time. This denotes an area for improvement and bridges the development gap between classification and regression.

Other challenges involve developing new relevant metrics to better describe the effectiveness of a model under continual learning strategy, including the concept of a model forgetting rate and the stability/plasticity related to the adaptation speed of a model’s loss function during its learning process. It would also be relevant to propose complementary measures related to the cost of continual learning strategies in terms of resource usage (CPU, GPU, memory, electricity) [56]. This would allow another kind of comparison with

standard incremental approach such as joint training. Also, it is currently recommended to provide solutions to make the model's results explainable and interpretable. This explainability serves the purpose of understanding how the model operates as well as providing legitimacy to the results obtained. If this was already investigated in the field of time series [57,58], continual learning becomes, at the same time, a new playground as well as a potential source of explainability to explore.

4. Concluding Remarks

The application of deep learning in the field of time series forecasting is still growing compared to common statistical approaches. One of the strengths of neural networks lies in their learning abilities. Although they are incremental in nature, their application is mostly static, with separate training and testing batches to validate their capacity. Their application is becoming increasingly ubiquitous in various domains of society, particularly in resource monitoring and demand prediction to effectively adapt any underlying process. Continual learning, by its name, is one of the techniques currently under development that are seen as key to address these issues by offering adaptability and stability solutions and ensuring the evolution and relevance of a service. The pioneering domain is for classification tasks, which currently receive much more attention than any other domains such as time series prediction. The three major approaches currently seen in the literature of continual learning focus on architecture, replay, and parameter regularization to meet expectations. These approaches provide the groundwork for building hybrid solutions more suited to each application case.

In recent years, there has been specific research applied on time series forecasting with continual learning strategies. One of the key points to consider in this domain, as in the others, is how patterns are reoccurring and how to avoid catastrophic forgetting. The other key point comes with the notion of tasks which are neither well defined nor managed in time series. In this domain, tasks are often implicit and we need task-free mechanisms and user interactions depending on application domains.

We also need to propose and agree on common benchmarks dedicated to continual learning. Similar to classification, this would allow for easy assessment of the performance of each model using multiple standard metrics including MSE, MAE, and RMSE, as well as domain-specific metrics such as, for example, the forgetting rate of a model or stability/plasticity metrics to define. It would also be relevant to propose measures related to the resource consumption (CPU, GPU, energy) needed over time for continual learning, in comparison with traditional methods. Analysis could be performed, considering the benefits in performances on one side compared to computational and energy resources involved. Budgeted strategies can also be explored.

Applications of continual learning are starting to emerge, especially in the fields of energy and economic forecasting for businesses. However, this remains relatively minor considering the numerous application domains where continual learning could have a real impact, such as in medicine and environmental studies (meteorology, climate, and hydrology) [10,27,59,60]. Further research and development are essential to unlock the full potential of continual learning in time series forecasting across diverse fields, paving the way for its broader adoption and integration into various real-world applications.

Author Contributions: Conceptualization, Q.B. and N.R.; methodology, Q.B.; software, Q.B.; validation, Q.B. and N.R.; formal analysis, N.R.; investigation, Q.B.; resources, Q.B.; data curation, Q.B.; writing—original draft preparation, Q.B.; writing—review and editing, N.R.; visualization, Q.B.; supervision, N.R.; project administration, N.R.; funding acquisition, N.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of the ARD JUNON project, granted by the Région Centre Val de Loire, FRANCE.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rusu, A.A.; Rabinowitz, N.C.; Desjardins, G.; Soyer, H.; Kirkpatrick, J.; Kavukcuoglu, K.; Pascanu, R.; Hadsell, R. Progressive Neural Networks. *arXiv* **2016**, 1606.04671.
2. van de Ven, G.M.; Tuytelaars, T.; Tolias, A.S. Three types of incremental learning. *Nat. Mach. Intell.* **2022**, *4*, 1185–1197. [[CrossRef](#)] [[PubMed](#)]
3. Parisi, G.I.; Kemker, R.; Part, J.L.; Kanan, C.; Wermter, S. Continual lifelong learning with neural networks: A review. *Neural Netw.* **2019**, *113*, 54–71. [[CrossRef](#)] [[PubMed](#)]
4. Gama, J.; Žliobaitė, I.; Bifet, A.; Pechenizkiy, M.; Bouchachia, A. A survey on concept drift adaptation. *ACM Comput. Surv.* **2014**, *46*, 1–37. [[CrossRef](#)]
5. 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*, 306–336. [[CrossRef](#)] [[PubMed](#)]
6. Baker, M.M.; New, A.; Aguilar-Simon, M.; Al-Halah, Z.; Arnold, S.M.; Ben-Iwhiwhu, E.; Brna, A.P.; Brooks, E.; Brown, R.C.; Daniels, Z.; et al. A domain-agnostic approach for characterization of lifelong learning systems. *Neural Netw.* **2023**, *160*, 274–296. [[CrossRef](#)] [[PubMed](#)]
7. Ao, S.I.; Fayek, H. Continual Deep Learning for Time Series Modeling. *Sensors* **2023**, *23*, 7167. [[CrossRef](#)]
8. Gunasekara, N.; Pfahringer, B.; Gomes, H.M.; Bifet, A. Survey on Online Streaming Continual Learning. *IJCAI Int. Jt. Conf. Artif. Intell.* **2023**, *2023*, 6628–6637.
9. Lange, M.D.; 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*, 3366–3385. [[PubMed](#)]
10. Hurtado, J.; Salvati, D.; Semola, R.; Bosio, M.; Lomonaco, V. Continual learning for predictive maintenance: Overview and challenges. *Intell. Syst. Appl.* **2023**, *19*, 200251. [[CrossRef](#)]
11. French, R.M. Catastrophic forgetting in connectionist networks. *Trends Cogn. Sci.* **1999**, *3*, 128–135. [[CrossRef](#)] [[PubMed](#)]
12. Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A.A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 3521–3526. [[CrossRef](#)] [[PubMed](#)]
13. Shin, H.; Lee, J.K.; Kim, J.; Kim, J. Continual Learning with Deep Generative Replay. In Proceedings of the NIPS 2017, Long Beach, CA, USA, 4–9 December 2017.
14. Li, Z.; Hoiem, D. Learning without Forgetting. In Proceedings of the ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016.
15. Zenke, F.; Poole, B.; Ganguli, S. Continual Learning Through Synaptic Intelligence. *PMC* **2017**, *70*, 3987–3995.
16. Buzzega, P.; Boschini, M.; Porrello, A.; Abati, D.; Calderara, S. Dark Experience for General Continual Learning: A Strong, Simple Baseline. *NeurIPS* **2020**, *33*, 15920–15930.
17. Lopez-Paz, D.; Ranzato, M. Gradient Episodic Memory for Continual Learning. In Proceedings of the NIPS 2017, Long Beach, CA, USA, 4–9 December 2017.
18. Aljundi, R.; Babiloni, F.; Elhoseiny, M.; Rohrbach, M.; Tuytelaars, T. Memory Aware Synapses: Learning what (not) to forget. In Proceedings of the ECCV 2018, Munich, Germany, 8–14 September 2018.
19. Chaudhry, A.; Ranzato, A.; Rohrbach, M.; Elhoseiny, M. Efficient Lifelong Learning with A-GEM. In Proceedings of the ICLR 2019, New Orleans, LA, USA, 6–9 May 2019.
20. Febrinanto, F.G.; Xia, F.; Moore, K.; Thapa, C.; Aggarwal, C. Graph Lifelong Learning: A Survey. *IEEE Comput. Intell. Mag.* **2022**, *18*, 32–51. [[CrossRef](#)]
21. Schwarz, J.; Luketina, J.; Czarnecki, W.M.; Grabska-Barwinska, A.; Teh, Y.W.; Pascanu, R.; Hadsell, R. Progress & Compress: A scalable framework for continual learning. In Proceedings of the International Conference on Machine Learning 2018, Vienna, Austria, 10–15 July 2018.
22. Chen, X.; Wang, J.; Xie, K. TrafficStream: A Streaming Traffic Flow Forecasting Framework Based on Graph Neural Networks and Continual Learning. In Proceedings of the IJCAI 2021, Montreal, QC, Canada, 19–26 August 2021.
23. Sokar, G.; Mocanu, D.C.; Pechenizkiy, M. Self-Attention Meta-Learner for Continual Learning. In Proceedings of the AAMAS 2021, Virtual Event, 3–7 May 2021.
24. 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. [[CrossRef](#)]
25. Rolnick, D.; Ahuja, A.; Schwarz, J.; Lillicrap, T.P.; Wayne, G. Experience Replay for Continual Learning. In Proceedings of the NeurIPS 2019, Vancouver, BC, Canada, 8–14 December 2019.
26. Smith, J.S.; 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 Workshops (CVPRW), New Orleans, LA, USA, 19–20 June 2022; pp. 2410–2420.

27. Hao, H.; Chu, Z.; Zhu, S.; Jiang, G.; Wang, Y.; Jiang, C.; Zhang, J.Y.; Jiang, W.; Xue, S.; Zhou, J. Continual Learning in Predictive Autoscaling. In Proceedings of the CIKM 2023, Birmingham, UK, 21–25 October 2023; pp. 4616–4622.
28. Bagus, B.; Gepperth, A. An Investigation of Replay-based Approaches for Continual Learning. In Proceedings of the IJCNN 2021, Shenzhen, China, 18–22 July 2021.
29. Grote-Ramm, W.; Lanuschny, D.; Lorenzen, F.; Brito, M.O.; Schönig, F. Continual learning for neural regression networks to cope with concept drift in industrial processes using convex optimisation. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105927. [[CrossRef](#)]
30. Yoon, J.; Yang, E.; Lee, J.; Hwang, S.J. Lifelong Learning with Dynamically Expandable Networks. In Proceedings of the ICLR 2017, Toulon, France, 24–26 April 2017.
31. Mirzadeh, S.I.; Chaudhry, A.; Yin, D.; Nguyen, T.; Pascanu, R.; Gorur, D.; Farajtabar, M. Architecture Matters in Continual Learning. *arXiv* **2022**, arXiv:2202.00275.
32. Hung, S.C.Y.; Tu, C.H.; Wu, C.E.; Chen, C.H.; Chan, Y.M.; Chen, C.S. Compacting, Picking and Growing for Unforgetting Continual Learning. In Proceedings of the NeurIPS 2019, Vancouver, BC, Canada, 8–14 December 2019.
33. Aich, A. Elastic Weight Consolidation (EWC): Nuts and Bolts. *arXiv* **2021**, arXiv:2105.04093.
34. Maschler, B.; Pham, T.T.H.; Weyrich, M. Regularization-based Continual Learning for Anomaly Detection in Discrete Manufacturing. *Procedia CIRP* **2021**, *104*, 452–457. [[CrossRef](#)]
35. He, Y.; Sick, B. CLear: An Adaptive Continual Learning Framework for Regression Tasks. *AI Perspect.* **2021**, *3*, 2. [[CrossRef](#)]
36. Pham, Q.; Liu, C.; Sahoo, D.; Hoi, S.C.H. Learning Fast and Slow for Online Time Series Forecasting. In Proceedings of the ICLR 2022, Baltimore, MD, USA, 25–29 April 2022.
37. Aljundi, R.; Lin, M.; Goujaud, B.; Bengio, Y. Gradient based sample selection for online continual learning. In Proceedings of the NeurIPS 2019, Vancouver, BC, Canada, 8–14 December 2019.
38. He, Y. Adaptive Explainable Continual Learning Framework for Regression Problems with Focus on Power Forecasts. *arXiv* **2021**, arXiv:2108.10781.
39. 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. [[CrossRef](#)]
40. 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. [[CrossRef](#)]
41. He, Y.; Henze, J.; Sick, B. Continuous learning of deep neural networks to improve forecasts for regional energy markets. *IFAC-PapersOnLine* **2020**, *53*, 12175–12182. [[CrossRef](#)]
42. Schillaci, G.; Schmidt, U.; Miranda, L. Prediction Error-Driven Memory Consolidation for Continual Learning: On the Case of Adaptive Greenhouse Models. *KI-Kunstl. Intell.* **2021**, *35*, 71–80. [[CrossRef](#)]
43. Gupta, V.; Narwariya, J.; Malhotra, P.; Vig, L.; Shroff, G. Continual Learning for Multivariate Time Series Tasks with Variable Input Dimensions. In Proceedings of the ICDM 2022, Orlando, FL, USA, 28 November–1 December 2022.
44. Farooq, J.; Bazaz, M.A. A deep learning algorithm for modeling and forecasting of COVID-19 in five worst affected states of India. *Alex. Eng. J.* **2021**, *60*, 587–596. [[CrossRef](#)]
45. Wang, H.; Li, M.; Yue, X. IncLSTM: Incremental Ensemble LSTM Model towards Time Series Data. *Comput. Electr. Eng.* **2021**, *92*, 107156. [[CrossRef](#)]
46. Fekri, M.N.; 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. [[CrossRef](#)]
47. Aljundi, R.; Kelchtermans, K.; Tuytelaars, T. Task-Free Continual Learning. In Proceedings of the NeurIPS 2018, Vancouver, BC, Canada, 3–8 December 2018.
48. Read, J.; Žliobaitė, I. Learning from Data Streams: An Overview and Update. *SSRN Electron. J.* **2022**. [[CrossRef](#)]
49. Vaswani, A.; Brain, G.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Łukasz Kaiser.; Polosukhin, I. Attention Is All You Need. In Proceedings of the NIPS 2017, Long Beach, CA, USA, 4–9 December 2017.
50. Gao, S.; Lei, Y. A new approach for crude oil price prediction based on stream learning. *Geosci. Front.* **2017**, *8*, 183–187. [[CrossRef](#)]
51. Zhang, Y.F.; Wen, Q.; Wang, X.; Chen, W.; Sun, L.; Zhang, Z.; Wang, L.; Jin, R.; Tan, T. OneNet: Enhancing Time Series Forecasting Models under Concept Drift by Online Ensembling. In Proceedings of the NeurIPS 2023, Vancouver, BC, Canada, 10–16 December 2023.
52. Melgar-García, L.; Gutiérrez-Avilés, D.; Rubio-Escudero, C.; Troncoso, A. Identifying novelties and anomalies for incremental learning in streaming time series forecasting. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106326. [[CrossRef](#)]
53. Zhao, L.; Kong, S.; Shen, Y. DoubleAdapt: A Meta-learning Approach to Incremental Learning for Stock Trend Forecasting. *Assoc. Comput. Mach.* **2023**, *8*, 3492–3503.
54. Sarmas, E.; Stropolas, S.; Marinakis, V.; Santori, F.; Bucarelli, M.A.; Doukas, H. An Incremental Learning Framework for Photovoltaic Production and Load Forecasting in Energy Microgrids. *Electronics* **2022**, *11*, 3962. [[CrossRef](#)]
55. Puah, B.K.; Chong, L.W.; Wong, Y.W.; Begam, K.M.; Khan, N.; Juman, M.A.; Rajkumar, R.K. A regression unsupervised incremental learning algorithm for solar irradiance prediction. *Renew. Energy* **2021**, *164*, 908–925. [[CrossRef](#)]
56. Prabhu, A.; Hammoud, H.A.A.K.; Dokania, P.; Torr, P.H.S.; Lim, S.N.; Ghanem, B.; Bibi, A. Computationally Budgeted Continual Learning: What Does Matter? In Proceedings of the CVPR 2023, Vancouver, BC, Canada, 18–22 June 2023.
57. Rojat, T.; Puget, R.; Filliat, D.; Ser, J.D.; Gelin, R.; Díaz-Rodríguez, N. Explainable Artificial Intelligence (XAI) on TimeSeries Data: A Survey. *arXiv* **2021**, arXiv:2104.00950.

58. Haque, S.; Eberhart, Z.; Bansal, A.; McMillan, C. Semantic Similarity Metrics for Evaluating Source Code Summarization. *IEEE Comput. Soc.* **2022**, *2022*, 36–47. [[CrossRef](#)]
59. Liu, B.; Xiao, X.; Stone, P. A Lifelong Learning Approach to Mobile Robot Navigation. *IEEE Robot. Autom. Lett.* **2020**, *6*, 1090–1096. [[CrossRef](#)]
60. Pal, G.; Hong, X.; Wang, Z.; Wu, H.; Li, G.; Atkinson, K. Lifelong Machine Learning and root cause analysis for large-scale cancer patient data. *J. Big Data* **2019**, *6*, 108. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.