International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 8, pp 12-21, August-2025

https://doi.org/10.47001/IRJIET/2025.907003

Lifelong Learning in Neural Networks: Techniques, Challenges, and Applications

Sura Saad Basher

Northern Technical University, Iraq

Abstract - **Lifelong Machine Learning (LML) is a versatile improvement to neural networks that in- creases models' ability to learn from sequential data in bite size and incrementally, while continually expanding generally acquired knowledge to new tasks. One of the persistent problems encountered in LML is known as catastrophic forgetting, whereby nets dislearn prior tasks upon exposure to new tasks. The following review explores these challenges in detail and presents fundamental neural network-based approaches to address such troubles in lifelong learning systems. In the edition, we prevent updates some of the key connectivist parameters while retaining prior knowledge from other tasks through regularization methods such as Elastic Weight Consolidation (EWC) and Learning without Forgetting (LwF). Even though useful, such strategies should be used with caution since they require as much emphasis on revisiting previous tasks as on acquiring new ones. Other rehearsal methods include the Partition Reservoir Sampling (PRS) and Optimizing Class Distribution in Memory (OCDM) that uses a portion of previous data for retraining, which can however prove rather space consuming for large-scale applications. Some architectural approaches, like the Compact, Picking, and Growing (CPG) principle, mean that the network structure grows with new tasks and extend from existing neurons or layers without influence from previous information. But these methods predetermine scalability since they increase computational complexity with the size of a casual network. Nevertheless, problems of how to deal with imbalance in data and shift in labels are still open problems particularly when applied in situations where the data distribution changes over time. However, lifelong learning in neural networks continue to experience growth challenges in catastrophic forgetting, scalability, and efficient knowledge transfer thus the need for further re- search. It will be crucial for applying neural networks for situations where it is required to learn over time but do not want to forget what has been learnt earlier.**

Keywords: Lifelong Learning, Neural Networks, Catastrophic Forgetting, Knowledge Transfer, Incremental Learning.

I. Introduction

Conventionally in the ML paradigm, a model is usually built from a given dataset to fit it into an algorithm. Cinem. Until this model is tested and the map that it will produce is representative of structures within the data, it is used. The above approach is aligned to the way previous ML algorithm work, where they are designed to solve a single problem or task [1]. On the other hand, Lifelong Machine Learning (LML) brings a concept of learning where different tasks are met in a continuous flow and each of them is related to a particular dataset [2]. In general, it is customary to oppose LML to Continual Learning (CL) in the ML research area, as both concepts present similar features; that is, why both terms are used interchangeably in the literature [3]. However, the problem of the demarcation between LML and CL still remains the subjects to discussions.

[2] That authors interested in deep learning, including deep neural networks in LML, generally refer to CL interchangeably with LML. Therefore, LML can be considered a more comprehensive notion under which CL can be viewed as a subfield dedicated to deep learning. This broader understanding of LML enables the investigations in works such as this one to encompass ample aspects of continual learning knowledge.

Humans have the inherent capability to learn and retrieve knowledge in one context constantly and transfer it to another. This capacity is essential for acquiring sensory-motor skills for storing as well as retrieving, long-term memories [4] In the context of ML, the idea of using prior knowledge when solving a new problem was considered which raises a question of how to use past experiences when sequentially mastering new problems. This is why what is referred to as knowledge transfer (KT) becomes crucial at this stage. KT is defined as the ability to transfer knowledge obtained from one task/context, to other tasks/contexts or environments [5]. KT in LML allows for the models to apply prior learnt knowledge to new tasks (forward KT) and to improve performance of the learnt tasks (backward KT).

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 9, Issue 8, pp 12-21, August-2025
https://doi.org/10.47001/IRJIET/2025.907003

II. Lifelong machine learning (LML)

Lifelong Machine Learning (LML) is an evolution of learning paradigm for continuously learning from new data streams presented in sequence wherein each stream is characterized by a specific task. More importantly, LML does not specify the total number of tasks that have to learned to meet the criterion [4]. Alphabetically, each should be learnable within the scope of individual learning, occurring in a system of multiple learning tasks in which each has individual aim and data. Figure 1 contrasts traditional ML with LML in terms of learning dynamics [6][2]. LML systems accumulate knowledge from previous tasks and adjust to new ones as they emerge. This adaptability, coupled with the ability to learn sequentially and incrementally, makes LML valuable for various research areas and practical applications [3]. When tasks are introduced in sequence, the process may involve expanding the class set, experiencing domain shifts in the input data, or an increase in the number of tasks. These factors are crucial for Knowledge Transfer (KT), as some techniques may not apply to every incremental learning scenario. As such, the processes of LML can be divided into three learning paradigms which corresponds to LML's step-by-step form, as well as incremental and continuous nature [7].

The first scenario, that is the Incremental Domain Learning (IDL), relates to the situation when the learner structure remains the same, while the input data distribution changes due to domain changes. In this scenario, the task identities (Task IDs or Task Boundaries) are not available during testing, and the output space remains fixed, with each task utilizing the same classes.

The second scenario, Incremental Class Learning (ICL), involves a learner progressively recognizing more classes by encountering tasks that introduce new classes over time. Here, task IDs are unnecessary, and the learner must classify all encountered classes. As new classes are introduced, the output dimension grows, allowing the model to integrate new classes into its knowledge base while generalizing across previously learned ones.

The third scenario, Incremental Task Learning (ITL), requires a learner to handle distinct tasks with different output spaces. Task IDs are known during both training and testing, enabling task- specific components like separate output heads or distinct models for each task. Output heads, which represent the final layers in a neural network model, are tailored for task-specific predictions. The goal of ITL is to leverage shared representations across tasks to enhance computational efficiency and facilitate knowledge transfer [8].

In this study, the term "task" is used consistently. In the context of IDL, a "task" refers to different domains introduced sequentially. For ICL, "task" refers to a set of classes introduced and learned collectively. In ITL, "task" refers to a distinct learning problem solved by the learner [4, 9]. This terminology aligns with the conventions used in recent literature. However, it is worth noting that some studies, including [7], use the term "context" to describe new domains in IDL and new class sets in ICL, while reserving the term "task" for ITL scenarios where task boundaries are known during both evaluation and training.

III. Lifelong Learning and Catastrophic Forgetting in Neural Networks

3.1 Lifelong Machine Learning

Lifelong learning represents a significant challenge in machine learning and neural network systems [10, 11] primarily due to catastrophic forgetting. This occurs when models, while learning new information, overwrite existing knowledge [12] . A lifelong learning system is an adaptive algorithm capable of learning continuously from a stream of information. The number of tasks to be learned, such as classification tasks, is not predefined, and new information must be integrated without forgetting or interference.

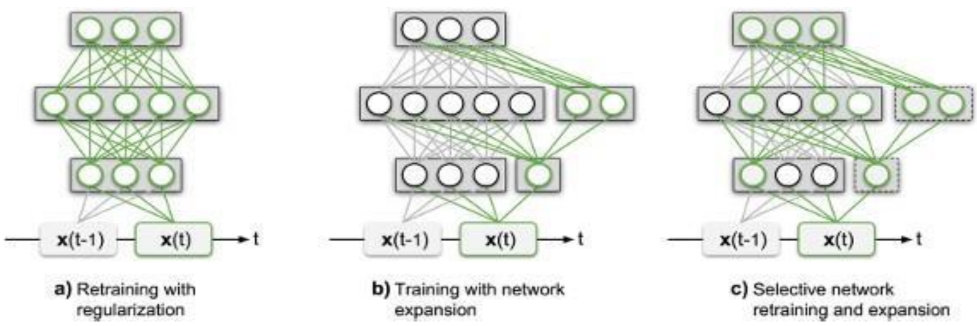

Figure 1: An illustration of neural network techniques for lifetime learning shows three strategies: (a) retraining with regularization to avoid catastrophic forgetting of previously learned tasks; (b) network extension with unchanging parameters to represent new tasks; and (c) selective retraining with potential expansion

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 8, pp 12-21, August-2025

https://doi.org/10.47001/IRJIET/2025.907003

In neural networks, catastrophic forgetting happens when new data differs substantially from previously encountered examples, leading the network to overwrite existing knowledge in shared representational resources [10, 13]. In traditional offline learning, this can be mitigated by repeatedly seeing the same examples, but this is not possible when data arrives in a continuous stream. As such, catastrophic forgetting has been extensively studied in neural networks trained through backpropagation [?, 14] and Hopfield networks [15].

Early solutions involved storing previous data and replaying it interleaved with new data to prevent forgetting [16, 17]. This method is still used [18, 19], but a drawback is the high memory requirement. Moreover, in cases with fixed neural resources, mechanisms must be designed to protect consolidated knowledge, such as the approaches by [20] and [21]. A promising solution is to allocate additional neural resources for new information [22, 23, 24], though this can lead to scalability issues. Predicting the number of tasks or samples in a lifelong learning scenario is difficult, making it non-trivial to allocate sufficient resources in advance [25].

Key strategies to avoid catastrophic forgetting include: (i) adding neural resources for new knowledge, (ii) using non-overlapping representations with fixed resources, and (iii) interleaving old and new information as tasks are introduced. Biological systems, through synaptic plasticity, have evolved to process continuous information without catastrophic forgetting [?, 26, 27, 21]. Unlike artificial systems, biological systems learn from dynamic, spatiotemporal input from the environment. In contrast, artificial neural networks are typically trained in batches with pseudo- randomized training data, which does not reflect the continuous learning nature of lifelong learning tasks

3.2 Regularization Approaches to Mitigate Catastrophic Forgetting

Regularization approaches aim to alleviate catastrophic forgetting by constraining updates to the neural network's weights. Inspired by theoretical neuroscience models, these approaches suggest that consolidated knowledge can be protected by synapses with varying plasticity [28, 29]. From a computational perspective, this is typically modeled by adding regularization terms to penalize changes in the neural network's mapping function. One well-known regularization approach is the Learning without Forgetting (LwF) method introduced by [30]. In LwF, a convolutional neural network (CNN) is used, and knowledge from previously learned tasks is transferred using knowledge distillation [31]. In this method, the new task's parameters, denoted as θ_n, are learned together with shared parameters θ_s, while ensuring the predictions from the old task's parameters θ_o do not deviate significantly from their original values. Given the new task's data (X_n, Y_n), old task predictions for the new data Y_o, and randomly initialized new parameters θ_n, the updated parameters $\theta_s{}^*$, $\theta_o{}^*$, $\theta_n{}^*$ are found by minimizing the loss:

$$(\theta s^*,\theta_o{}^*,\theta_n{}^*)\leftarrow \arg\min_{\theta s,\theta o,\theta n} \lambda o L_{\mathrm{old}}(Yo,\hat{Y}o)+L_{\mathrm{new}}(Yn,\hat{Y}n)+R(\theta s,\theta o,\theta n)$$

where $L_{\mathrm{old}}(Yo,\ \hat{Y}o)$ and $L_{\mathrm{new}}(Yn,\ \hat{Y}n)$ are the loss functions for the old and new tasks, respectively, and R is a regularization term that prevents overfitting. A balancing factor λo is used to weigh the importance of old versus new tasks. Although this method helps prevent catastrophic forgetting, it has limitations, such as the increasing training time with the addition of new tasks. Another important method is Elastic Weight Consolidation (EWC), proposed by [20], which applies a quadratic penalty on changes to task-relevant weights. The relevance of parameters θ with respect to task data D is modeled by the posterior distribution $p(\theta|D)$. Assuming independent tasks A and B, the posterior probability is given by Bayes' rule:

$$\log p(\theta|D) = \log p(D_B|\theta) + \log p(\theta|D_A) - \log p(D_B)$$

The loss function for EWC is then approximated as:

$$L(\vartheta) = L_B(\vartheta) + \frac{\lambda}{2}\sum_i F_i(\vartheta_i - \vartheta^*_{A,i})^2$$

Where L_B is the loss of task B, λ is a scaling factor, and F_i is the Fisher information matrix. This method slows down learning for parameters that are important for previous tasks, but it may not perform well when incrementally learning new categories.

In summary, regularization methods offer strategies to mitigate catastrophic forgetting by protecting important neural weights. However, these approaches often introduce trade-offs in performance between old and new tasks, especially when neural resources are limited.

IV. Literature Review

Roseberry et al. introduced a self-adapting algorithm designed to learn from multi-label drifting data streams, addressing both the challenges of continuous learning and adapting to concept drifts [32]. Their research emphasizes the additional complexities that arise when learning from multi-label data, particularly when the statistical properties of the data change over time due to drift. Traditional learning models often struggle to maintain accuracy as data distributions evolve. To tackle this issue, the authors propose an approach that leverages the k-Nearest Neighbors (kNN) algorithm,

ISSN (online): 2581-3048
Volume 9, Issue 8, pp 12-21, August-2025
https://doi.org/10.47001/IRJIET/2025.907003

known for its simplicity and effectiveness in classification tasks. The kNN algorithm bases its predictions on the proximity of the test instance to the k nearest examples in the training set. However, unlike the conventional kNN, this method incorporates a self-adjusting mechanism that dynamically modifies the value of k based on the characteristics of the incoming data, enabling the model to better respond to concept drift and changing label distributions in dynamic environments.

Additionally, while traditional kNN uses majority voting to assign classes, this approach encounters difficulties in multi-label contexts, where each neighbor may have a distinct set of labels. To address this, the algorithm calculates the frequency of each label's occurrence among the closest neighbors and computes the likelihood and posterior probability of each label using Bayesian principles. The decision on whether a label is present or absent is based on these computed probabilities.

Masuyama et al. combine Adaptive Resonance Theory (ART) [33] with a Bayesian approach for label probability computation to cluster instances with similar label patterns, reducing the complexity of the output space to be learned [34]. ART, a cognitive information processing theory, is used to support neural network models by addressing the stability-plasticity dilemma through competitive and self-organizing processes. The algorithm which has been introduced by Masuyama et al., pertains to ART-based algorithm where alternatives are generated in the form of nodes, or classifiers that learn and improve on their output from the arrival of new data. At the same time, the Bayesian method records the frequencies of each label separately and computes the probability in the same way as Roseberry et al. [32]. These probabilities are used to jointly label the instances with each label represented by a presence which is threshold at 0.5 while the absence is below the threshold value.

The MLCA algorithm, developed from this work, appears effective in adapting to and being scalable upon suitably designed problems in multi-label learning. It is capable of continuous learning and the classification performance on both synthetic and real datasets of the proposed method is competitive. As for the partial label image recognition, different from the standard partial labels, when a model is trained for one partial label image recognition task and then for another, it is likely to forget the previous learned features which-the Augmented Graph Convolutional Network (AGCN) approach [35], [36] designed to solve. AGCN then constructs the ACM across consecutive partial-label tasks in order to capture label dependencies in association with a dynamic matrix structure that can provide better label representation. This allows knowledge transfer and tailoring when it comes to lifelong learning situations. Compared with the existing methods that train the model only with multi-label images and without considering the relationships between different labels or images, AGCN uses graph convolutional networks to capture label dependencies and learns the relationship-preserving loss function, which makes multi-label recognition more efficient and accurate.

Continual learning is discussed in [37] where Wang et al. propose a few-shot continual learning framework for audio classification. Few-shot learning identifies new classes over a small number of examples at inference time, which enable quick model updates. The framework is grounded on the Dynamic Few-Shot Learning (DFSL) paradigm which it uses a CNN classifier for feature extraction, and an attention mechanism for the use of past experiences. To this end, while using this model, the authors chose to employ categorical cross-entropy for the multi-label classification rather than the traditional approach of using binary cross-entropy. This shift means that each label is treated separately as binary classification problem which makes it easier to train the model and improve on its ability to detect multiple labels at once.

Kim et al posed catastrophe forgetting in learning patterns of neuRal networks from imbalanced datasets with partition reservoir sampling or PRS [38]. This is an extension of the Reservoir Sampling, the common method of storage and sampling of stream data. For this reason, PRS modifies the approach through coming up with balanced training partitions in that mini batches should have both the majority and the minority classes. It also helps to overcome the problems of class imbalance during the lifelong learning with the help of imbalanced data.

Similarly, Optimizing Class Distribution in Memory (OCDM) [39] is a memory based method which maintains a representative class distribution in memory. In OCDM, memory updates are presented as an optimization problem with the goal of maintaining a vicinity of a target distribution of samples in memory, usually quantified by the distance between the current and target cumulative distribution functions (CDFs) using a measure such as the Kullback-Leibler (KL) divergence.

Pham et al. propose a lifelong topic modeling method to uncover hidden topics in text corpora by leveraging prior domain knowledge [40]. This approach uses a probability-based close domain metric to select valuable knowledge from past tasks to improve current learning. The method enhances multi-label text classification by deriving hidden topics from domain closeness, and it employs a multi-label learning algorithm [41] that first clusters positive and negative

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 8, pp 12-21, August-2025

https://doi.org/10.47001/IRJIET/2025.907003

instances for each label and then develops label-specific features, followed by a binary classifier for each label.

BAT-OCDM [42] focuses on Domain Incremental Learning, where the set of output labels remains unchanged across tasks, but input data distribution changes from task to task. The approach modifies the OCDM algorithm [39] by employing separate memory for each task to balance labels and tasks, ensuring that previous tasks are not forgotten. This method was tested on real-world data from the packaging industry to predict alarms, modeled as a multi-label classification problem. Chen et al. explore deep lifelong learning for defect detection in manufacturing pipelines [43].

Their method uses densely connected neural networks, allowing the model to detect both new and previously known defect types without retraining on old data. Each task is converted into a binary classification task, and new tasks are learned with the help of CPG algorithm. It has been observed that the CPG algorithm finds out important weights, makes the model lighter and in case of need increases the size of the network.

Last, the Knowledge Restore and Transfer (KRT) framework [44] is proposed for MLCIL to enhance the stability of learned representations under distribution shifts. It incorporates two modules: The other two suggested components are the Dynamic Pseudo-Label (DPL) module, which reconstructs the knowledge from the previous labels, and the Incremental Cross-Attention (ICA) module, which maintains the specific task knowledge and passes it to the new architecture. This method enhances the recognition performance and handles the problem to do with forgetting in MLCIL tasks.

Table 1: Summary of algorithms, lifelong learning techniques, multi-label methods, datasets, and references from the literature review

Algorithm	LL Technique/ Method	Multi-label Technique/ Method	Base Algorithm	Datasets	Purpose	Ref
AGCN	Augmented Correlation Matrix	Graph Convolutional Network (GCN)	GCN	COCO, VOC	Multi-label Image recognition with partial labels	[35], [36]
CIFDM	Knowledge Compression, Knowledge Bank, Pioneer Module	Dynamic Label assignment	Feature distillation	Multi-label Image stream datasets	Continuous Learning and catastrophic forgetting prevention	[45]
CDSH	Semantic Hashing	Binary code Learning with increasing labels	Hashing Networks	Image datasets	Learning Binary codes for multi- label images	[46]
DLFL	Feature Disentanglement	One-Specific-Feature-for-One-Label (OFOL)	Cross-Attention Mechanism	Multi-label classification datasets	Disentangled Label feature learning	[47]
KRT	Knowledge Restore and Transfer (KRT)	Dynamic Pseudo- Label (DPL), Incremental Cross-Attention (ICA)	Neural Networks	Multi-label incremental datasets	Mitigating catastrophic forgetting inclass-incremental tasks	[44]
OCDM	Memory-based rehearsal	Balanced Class distribution	Neural Networks	Imbalanced datasets	Handling imbalanced data during lifelong learning	[39]

V. Techniques and Challenges in Lifelong Learning for Neural Networks

Lifelong learning in neural networks has become a critical area of research, driven by the need to continuously adapt to new tasks and data without suffering from catastrophic forgetting. To address this challenge, various techniques have been developed to enable neural networks to retain knowledge from previously learned tasks while efficiently learning new information. Despite significant

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 8, pp 12-21, August-2025

https://doi.org/10.47001/IRJIET/2025.907003

advancements, lifelong learning still faces several challenges, which are discussed alongside the primary techniques used to mitigate catastrophic forgetting.

Regularization approaches are among the most commonly used techniques for lifelong learning. These methods, such as Elastic Weight Consolidation (EWC) and Learning without Forgetting (LwF), work by constraining weight updates to prevent the overwriting of previously learned knowledge. Additional terms are introduced into the loss function to penalize significant changes to parameters that are crucial for older tasks. While effective, regularization techniques require careful tuning to balance the need to retain past knowledge with learning new information.

Another method is rehearsal, which involves storing a small subset of previously seen data and revisiting these examples when learning new tasks. This ensures that the network can be trained on a mixture of old and new data, improving retention of previously learned tasks. Techniques like Partition Reservoir Sampling (PRS) and Optimizing Class Distribution in Memory (OCDM) rely on this approach, balancing the memory buffer to mitigate class imbalance and ensure continuous learning without catastrophic forgetting.

Architectural approaches focus on dynamically expanding the network to accommodate new tasks. Methods such as Compact, Picking, and Growing (CPG) introduce new neurons or subnetworks for new tasks while preserving the existing model structure for older tasks. This strategy allows the network to continue learning without overwriting previously acquired knowledge, but it may lead to scalability issues due to increased resource demands.

However, lifelong learning has its own drawbacks which people experience even today. Another problem which remains prevalent is called catastrophic forgetting, whereby a model forgets previously acquired knowledge upon exposure to new tasks. While it does get somewhat lessened with the use of regularization and rehearsal methods, this remains a difficulty, especially in cases where the new task specifically differs markedly from prior scenarios.

Another is the problems of scale and resource on which many effective lifelong learning techniques depend. The architectonic approaches that extend the exposition of a new task aggravate computational and memory costs, and hence, non-scalability. As with the rehearsal methods, there may be definite difficulties linked to memory capacity, when loads of previous information are to be stored in order for the proper balance for learning can be maintained.

The new materials with shifted class distributions also cause problems with data that is imbalanced and lifelong learning. Methods like PRS and OCDM solve these problems by varying the data stored in the rehearsal buffer but the problem of in which the strategies for rehearsal buffer reflect optimal performance is still open.

Another valuable yet very challenging aspect is when information is to be passed between tasks. Continuous learning is a double process in which knowledge acquired in one task enhances the performance in another task and the knowledge acquired in a new task improves performance on an older task. Nevertheless, the problem of balancing these transfers without the negative consequences, including over-[U+FB01]tting and knowledge erosion, is challenging.

In conclusion, various approaches have been proposed to retain lifelong learning in neural network which yet include difficulties like catastrophic forgotten, scalability, imbalanced data and efficient transfer of knowledge. More work is being conducted to improve these techniques and to develop others that will enable more effective lifelong learning architecture to be put in place.

VI. Discussion

Lifelong machine learning (LML) and the idea of applying it to neural networks have attracted increasing interest in recent years especially in cases when it is needed to learn new tasks without losing the previous knowledge. This section also looks into broader possibilities of lifelong learning, probable uses and the existing challenges that call for further development of this concept.

One of the most important aspects of the LML is the possibility to model the human learning process step-by-step or in a From Low to Metamel form. In practice, data arrives incrementally and sequentially and models need to learn how to handle changing data without repeating training. For example, in scenarios involving change as in autonomous cars, wellbeing technology and robotics, LLM lets models to learn new information in the subsequent task and at the same time, preserve significant information gleaned from prior tasks. Lack of catastrophic forgetting is important in such applications as it is necessary to be able to learn incrementally in environments important for safety, precision, and time constant.

However, it has to be noted that while considerable progress has been made in taming catastrophic forgetting, few problems still remain unsolved. At the same time, the discussed approaches that include the regularization approaches, rehearsal methods, and architectural modifications, are not free from the trade-offs. There are some forms such as Elastic Weight Consolidation (EWC) that can minimize catastrophic forgetting, yet they demand additional

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 8, pp 12-21, August-2025

https://doi.org/10.47001/IRJIET/2025.907003

fine-tuning for work and very limited learning for new work if the constraints are tight. Like it, rehearsal methods also work well in revision by causing activation of old memories and thus calls for memory storage of past data that may not always be possible, particularly in areas where issues to do with privacy or availability of resources will come into play.

It is also important to note that the issue of LML techniques is difficult to scale up. Since models are expected to learn a number of tasks as the years progress, the computational and memory loads increase, especially with architectural methods that widen the network. This brings about the question of the feasibility of implementing these techniques in real-world applications of large scale. The future of this area of work lies in identifying how it is possible to achieve and maximize lifelong adaptability while minimizing model complexity.

Additionally, we have the problem of data imbalance and label shifts as the next challenge. In multitask learning, new tasks or data may bring considerable changes in the distribution of labels or feature space, which may translate into changes in cross-task generalization performance.

To address this, methods such as Partition Reservoir Sampling (PRS) and Optimizing Class Distribution in Memory (OCDM) try to maintain balanced training data in memory but still need improvements that work well for highly skew or rare classes.

They also found that the concept of knowledge transfer is also important in lifelong learning. While forward knowledge transfer, in which new knowledge acquired facilitates learning of other new tasks, has been investigated, previous studies have paid little attention to how new information could enhance the performance of previously learned tasks, which is known as backward knowledge transfer. Locating the best ways in which both kinds of transfer may be accomplished without leading to detrimental consequences like over-learning or interference from, or to, other tasks has however been viewed as a major research issue.

However, there are other challenges that are unethical and societal in nature which affect the deployment of the lifelong learning systems. The constant update of models leads to the issue of decision-making in AI concerning decision sustainability and interpretability. Because models change over time and new information may be introduced in their construction, it is crucial to maintain interpretability of the outcomes and actions of the models throughout that process of utilization, especially when the decisions are made in critical areas such as medicine or vehicle driving. In addition, the require a strong safeguards preventing models learned with such data from reinforcing bias or transmitting it as the model works on new data sources.

In conclusion, lifelong learning offers promising marginal benefits for developing even more adaptable, efficient, and intelligent organizations. Despite these achievements, there are emergent issues on catastrophic forgetting, scalability, and imbalanced data which requires a further study and development of sophisticated, more effective approaches. Responsible lifelong learning systems also have to be designed with approaches that advance not only technical criteria, but also the system's interpretability and ability to remain or become fair throughout the system's life cycle. This field remains promising, although work will need to progress to continue the development of residual learning for neural networks for lifelong learning.

VII. Conclusion

Lifelong Machine Learning (LML) is a novel framework in Machine learning, which aims to support models that are able to learn perpetually and in parallel accomplish new tasks but never forget what they have previously learned. Consequently, LML systems provide significant importance in different practical applications of real life as; LML systems are very flexible and can be scaled up or scaled down, depending on the application domain, thus these are very important in special applications such as in healthcare management, self-driving cars, and robotics which requires continuous learning.

In this work, we looked into several ways to reduce catastrophic forgetting which includes regularization, rehearsal methods, and architectural modifications. Both approaches come with unique benefits, while regularization techniques help maintain existing knowledge, rehearsal approaches are practical in the context of selective memory replay. However, each technique also comes with some disadvantages especially related to scalability, forgotten task memory optimization and the trade-off between old and the new task performance. Nonetheless, fundamental developments in forward and backward knowledge transfer suggest great potential for enhancing learning within models.

Although advances in developing LML theories have marked considerable progress, the improvement of this theoretical approach will not be smooth and is accompanied by several issues that require further investigation. These are issues such as the handling of imbalancing data and label, the scalability of the lifelong learning system, and the strategies for transferring knowledge among tasks. Furthermore, as LML systems are adopted more commonly, it will be imperative that the system functions in an ethically correct manner and be able to also maintain transparency, fairness and explainability

ISSN (online): 2581-3048
Volume 9, Issue 8, pp 12-21, August-2025
https://doi.org/10.47001/IRJIET/2025.907003

in the continuous learning loops. Therefore, the capability of lifelong learning in machine learning and neural networks is thus big. If the current remaining challenges are tackled and current techniques are further developed, lifelong learning systems can transform into very efficient and scalable solutions to solve long and complicated real-life issues. Subsequent development and exploration of interests in this particular field of research will greatly help in achieving enhanced adaptation, optimization, and intelligent performance of LML and the AI systems that will arise from it in the future.

REFERENCES

[1] H. Liu and M. Cocea. Traditional machine learning. In Machine Learning, pages 11–22. *Springer International Publishing*, Cham, 2018.

[2] Z. Chen and B. Liu. Lifelong machine learning, 2018.

[3] Z. Mai, R. Li, J. Jeong, and et al. Online continual learning in image classification: An empirical survey. *Neurocomputing*, 469:28–51, 2022.

[4] G. I. Parisi, R. Kemker, J. L. Part, and et al. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019.

[5] D. Kudithipudi, M. Aguilar-Simon, J. Babb, and et al. Biological underpinnings for lifelong learning machines. *Nature Machine Intelligence*, 4(3):196–210, 2022.

[6] X. Wang, Y. Chen, and W. Zhu. A survey on curriculum learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4555–4576, 2022.

[7] G. M. Van de Ven, T. Tuytelaars, and A. S. Tolias. Three types of incremental learning. *Nature Machine Intelligence*, 4(12):1185–1197, 2022.

[8] Y. C. Hsu, Y. C. Liu, A. Ramasamy, and et al. Re-evaluating continual learning scenarios: A categorization and case for strong baselines. *arXiv preprint arXiv:*1810.12488, 2018.

[9] M. De Lange, R. Aljundi, M. Masana, and et al. A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3366–3385, 2022.

[10] R. M. French. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3(4):128–135, 1999.

[11] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick. Neuroscience-inspired artificial intelligence. *Neuron Review*, 95(2):245–258, 2017.

[12] S. Thrun and T. Mitchell. Lifelong robot learning. *Robotics and Autonomous Systems*, 15:25–46, 1995.

[13] M. McCloskey and N. J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. *The Psychology of Learning & Motivation*, 24:104–169, 1989.

[14] R. Ratcliff. Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review*, 97(2):285–308, 1990.

[15] J. P. Nadal, G. Toulouse, J. P. Changeux, and S. Dehaene. Networks of formal neurons and memory palimpsets. *Europhysics Letters*, 1:535–543, 1986.

[16] A.V. Robins. Catastrophic forgetting in neural networks: The role of rehearsal mechanisms. *In Proceedings of the First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems. IEEE Computer Society Press*, 1993.

[17] A.V. Robins. Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science*, 7(2):123–146, 1995.

[18] A.Gepperth and C. Karaoguz. A bio-inspired incremental learning architecture for applied perceptual problems. *Cognitive Computation*, 8(5):924–934, 2015.

[19] S. A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert. icarl: Incremental classifier and representation learning. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[20] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, and et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

[21] F. Zenke, B. Poole, and S. Ganguli. Continual learning through synaptic intelligence. *In Proceedings of the International Conference on Machine Learning (ICML), Sydney, Australia*, 2017.

[22] J. Hertz, A. Krogh, and R. G. Palmer. Introduction to the Theory of Neural Computation. *Addison-Wesley, Redwood City, CA*, 1991.

[23] G. I. Parisi, J. Tani, C. Weber, and S. Wermter. Lifelong learning of humans actions with deep neural network self-organization. *Neural Networks*, 96:137–149, 2017.

[24] A.A. Rusu, N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, and et al. Progressive neural networks. *In Proceedings of the International Conference on Learning Representations (ICLR)*, 2016.

[25] F. M. Richardson and M. S. C. Thomas. Critical periods and catastrophic interference effects in the development of self-organising feature maps. *Developmental Science*, 11(3):371–389, 2008.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 8, pp 12-21, August-2025

https://doi.org/10.47001/IRJIET/2025.907003

[26] M. M. Murray, D. J. Lewkowicz, A. Amedi, and M. T. Wallace. Multisensory processes: A balancing act across the lifespan. *Trends in Neurosciences*, 39:567–579, 2016.

[27] J. D. Power and B. L. Schlaggar. Neural plasticity across the lifespan. Wiley Interdisciplinary Reviews: *Developmental Biology*, 6(216), 2016.

[28] M. K. Benna and S. Fusi. Computational principles of synaptic memory consolidation. *Nature Neuroscience*, 19(12):1697–1708, 2016.

[29] S. Fusi, P. J. Drew, and L. F. Abbott. Cascade models of synaptically stored memories. *Neuron*, 45(4):599–611, 2005.

[30] Z. Li and D. Hoiem. Learning without forgetting. *In Proceedings of the European Conference on Computer Vision (ECCV)*, pages 614–629, Amsterdam, The Netherlands, 2016.

[31] G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. *In Advances in Neural Information Processing Systems (NIPS), Workshop on Deep Learning and Representation*, 2014.

[32] M. Roseberry, B. Krawczyk, Y. Djenouri, and A. Cano. Self-adjusting k nearest neighbors for continual learning from multi-label drifting data streams. *Neurocomputing*, 442:10–25, Jun 2021.

[33] G. A. Carpenter and S. Grossberg. Adaptive resonance theory. Technical Report Tech. Rep. CAS/CNS TR-98-029, *Dept. Cognitive Neural Systems, Center for Adaptive Systems, Boston University, Boston, MA*, USA, 2010.

[34] N. Masuyama, Y. Nojima, C. K. Loo, and H. Ishibuchi. Multi-label classification via adaptive resonance theory-based clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8696–8712, Jul 2023.

[35] K. Du, F. Lyu, F. Hu, L. Li, W. Feng, F. Xu, and Q. Fu. Agcn: Augmented graph convolutional network for lifelong multi-label image recognition. *In Proceedings of the IEEE International Conference on Multimedia Expo (ICME)*, pages 1–6, Jul 2022.

[36] K. Du, L. Li, F. Lyu, F. Hu, Z. Xia, and F. Xu. Class-incremental lifelong learning in multilabel classification. 2022. arXiv:2207.07840.

[37] Y. Wang, N. J. Bryan, M. Cartwright, J. P. Bello, and J. Salamon. Few shot continual learning for audio classification. *In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 321–325, Jun 2021.

[38] C. D. Kim, J. Jeong, and G. Kim. Imbalanced continual learning with partitioning reservoir sampling. In Computer Vision—ECCV, pages 411–428, Glasgow, U.K., 2020. *Springer*.

[39] Y.-S. Liang and W.-J. Li. Optimizing class distribution in memory for multi-label online continual learning. 2022. arXiv:2209.11469.

[40] T.-N. Pham, Q.-T. Ha, M.-C. Nguyen, and T.-T. Nguyen. A probability based close domain metric in lifelong learning for multi-label classification. *In Advanced Computational Methods for Knowledge Engineering*, pages 143–149, Cham, Switzerland, 2020. *Springer*.

[41] M.-L. Zhang and L. Wu. Lift: Multi-label learning with label-specific features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1):107–120, Jan 2015.

[42] D. Dalle Pezze, D. Deronjic, C. Masiero, D. Tosato, A. Beghi, and G. A. Susto. A multilabel continual learning framework to scale deep learning approaches for packaging equipment monitoring. *Engineering Applications of Artificial Intelligence*, 124, Sep 2023.

[43] C.-H. Chen, C.-H. Tu, J.-D. Li, and C.-S. Chen. Defect detection using deep lifelong learning. *In Proceedings of the IEEE 19th International Conference on Industrial Informatics (INDIN)*, pages 1–6, Jul 2021.

[44] S. Dong, H. Luo, Y. He, X. Wei, and Y. Gong. Knowledge restore and transfer for multi-label class-incremental learning. 2023. *arXiv:*2302.13334.

[45] Y. Wang, Z. Wang, Y. Lin, L. Khan, and D. Li. Cifdm: Continual and interactive feature distillation for multi-label stream learning. *In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2121–2125, Jul 2021.

[46] G. Song, K. Huang, H. Su, F. Song, and M. Yang. Deep continual hashing for real-world multi-label image retrieval. *Computers Vision and Image Understanding*, 234, Sep 2023.

[47] J. Jia, F. He, N. Gao, X. Chen, and K. Huang. Learning disentangled label representations for multi-label classification. 2022. *arXiv:*2212.01461.

ISSN (online): 2581-3048
Volume 9, Issue 8, pp 12-21, August-2025
https://doi.org/10.47001/IRJIET/2025.907003

Citation of this Article:

Sura Saad Basher. (2025). Lifelong Learning in Neural Networks: Techniques, Challenges, and Applications. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(8), 12-21. Article DOI https://doi.org/10.47001/IRJIET/2025.908003

****

