ISSN: 2182-2069 / E-ISSN: 2182-2077

Enhancing Continual Learning in Neural Networks with Lightweight Beneficial Perturbation Networks

Vanaparthi Kiranmai[1*], and Dr.A. Manikandan[2]

[1*]Ph.D Research Scholar, Department of Computer Science and Engineering (CSÈ), Vels University, Palavaram, Chennai, Tamil Nadu, India. kiranmaibolishetti18@gmail.com, https://orcid.org/0009-0003-9330-2189

[2]Research Supervisor, Vels University, Palavaram, Chennai, Tamil Nadu, India. mani.se@vistas.ac.in, https://orcid.org/0009-0000-9483-3537

Received: September 18, 2025; Revised: October 24, 2025; Accepted: December 19, 2025; Published: February 27, 2026

Abstract

The artificial intelligence systems are bombarded by constant learning, and adaptive algorithms are necessary to sustain knowledge across a broad spectrum of tasks. The paper discusses the performance of various techniques for handling incremental and object recognition problems using the CIFAR-100 dataset. We present new algorithms, Beneficial Perturbation Network (BPN) variants, namely BD+EWC, PSP, and BD+PSP, aimed at improving flexibility and resiliency in ongoing learning. We do this in our study by comparing the algorithms based on their accuracy, performance, and cost per calculation task. Results show that the highest accuracy of 90.65 is with BD+PSP, followed by PSP with 90.01, and lastly, BD+EWC with 89.95. To promote safe data processing, we will incorporate encryption-based processes and privacy-preserving algorithms into the proposed models to ensure that sensitive data is not destroyed during the learning process. In addition, cost analysis proves that BD+EWC enjoys the minimal computational overhead per task of 4,039 bytes, and PSP and BD+PSP have 10,897 bytes and 11,456 bytes, respectively. The paper indicates the potential of safe and efficient systems of continuous learning in real-life contexts, particularly where accuracy and confidentiality of data are needed. Moreover, we analyze the performance patterns of different variants on incremental CIFAR-100 tasks. The most effective strategies, including BD+PSP, display higher retention of learned information on succession tasks. Our proposed algorithms highlight the immense advances in object recognition tasks—and the algorithms themselves—when compared to the current methods. The results underscore the algorithms' importance in object recognition as they achieve results superlative to the existing ones. LW-BPN-EPIE-Net's architecture is optimized and experimental results show it is helpful in the accurate estimation of BPM as well as its ability to function in continuous learning systems. Such systems as LW-BPN-EPIE-Net demonstrate its ability to maintain an accuracy of 98 percent. The model achieves 6% on CIFAR-100, outperforming existing models, and its processing is efficient, taking only 28 seconds. Moreover, we go further to evaluate the validity of current empirical Beneficial Perturbation Networks and processes to understand the complexity of issues encountered by the detection departments. Our research goal, therefore, is to develop a lightweight yet comprehensive framework to address the problem of constant learning in neural networks. Our observations provide helpful advice for creating adaptive algorithms capable of performing various and changing tasks in practice.

Journal of Internet Services and Information Security (JISIS), volume: 16, number: 1 (February-2026), pp. 253-276.
DOI: *10.58346/JISIS.2026.I1.015*

*Corresponding author: Ph.D Research Scholar, Department of Computer Science and Engineering (CSÈ), Vels University, Palavaram, Chennai, Tamil Nadu, India.

Keywords: Neural Networks, Adaptive Algorithms, Beneficial Perturbation Network, Lightweight, CIFAR-100 Tasks, Security in Machine Learning.

1 Introduction

The innovation of artificial intelligence, particularly in continuous learning, is Optimizing Continuous Learning in Neural Networks using Lightweight Beneficial Perturbation Networks (LW-BPN) (Wang et al., 2023). As AI systems are increasingly implemented in practice, the security and privacy of neural network models have taken center stage, particularly in internet-based applications. Perpetual learning or continuous learning is a process required of artificial intelligence systems so that they can adapt and evolve when new tasks and information are introduced to them without forgetting what they already learned (Mundt et al., 2023). The adoption of advanced algorithms capable of continuous learning within the scope of cybersecurity policies and internet services should also be geared toward addressing adversarial attacks, data breaches, and model inversion while maintaining data confidentiality. Still, traditional artificial neural networks tend to forget operations due to their architecture and training history, which is one of their many challenges (Duncker et al., 2020). In response to these challenges, the LW-BPN approach implements secure federated learning and encryption protocols which ensures that learned models do not reveal sensitive data. The authors propose the Lightweight Beneficial Perturbation Network, which addresses the catastrophic forgetting problem in neural networks (Gupta et al., 2024). Furthermore, trust and belief in the data accessed for continuous learning are crucial, especially for safeguarding sensitive personal or confidential data. Pertinent data to the models must be accurate. LW-BPN's main idea is to infuse functional disturbances into the training of neural nets, teaching them new objectives and retaining the mastered ones (Ge et al., 2023). To protect the user, differential privacy and homomorphic encryption are employed to ensure that the personal information is irretrievably encoded, while still enabling the proper training of the model under concern. Beneficial training perturbation can be classified as subtle shifts in the neural architecture that remain within the core of the learned objectives (Sun et al., 2022). The network's ability to focus on essential aspects and avoid overfitting them to specific samples is strengthened by these perturbations (Chowdhury et al., 2022). LW-BPN strives to maintain equilibrium between new insights and previously gained knowledge by carefully perturbing these boundaries (Ge et al., 2023). For LW-BPN to be lightweight, flexible, and secure, it employs privacy-by-design approaches such as federated learning and encryption, which guarantee that sensitive data is protected during the learning processes. Such measures enhance not only the model's defense against data exfiltration and adversarial threats but also the model's integrity during multi-task learning (Du et al., 2023). LW-BPN's defensed training data makes it highly suitable for deployment in pragmatic environments where utmost privacy is required, such as in healthcare, finance, and critical infrastructure (Huang et al., 2017).

Another essential characteristic of LW-BPN is its light weight, as it simplifies scaling to large topologies in neural networks without sacrificing computing efficiency (Fedorin et al., 2022). Unlike many other methods of continuous learning, LW-BPN's focus on the present does not require extensive historical data, nested constructs, or computational architecture for resource-intensive real-world applications (Vijeyakumar et al., 2018). With respect to neural networks, LW-BPN has been tested on a wide range of neural network designs and benchmark datasets (Fedorin et al., 2022). The results show that LW-BPN outperforms other continuous learning implementations in accuracy and memory efficiency (Narayana et al., 2022). LW-BPN also possesses a high degree of tolerance in the face of different forms of learning, such as sequential and interspersed task learning, which is an indicator of its versatility and usefulness (Narayana et al., 2022). Additionally, the authors elaborated on the mechanisms of LW-BPN and its impact on the learning characteristics of neural networks (Wang et al., 2022). They explained how LW-BPN aids neural networks in achieving a balance between forgetting and stability when learning new tasks and retaining old information, particularly in the study of learned representations and patterns of perturbation (Cheng et al., 2020). In addition to the empirical proof, LW-BPN is also theoretically examined in relation to aspects of cognitive science and neuroscience. The constellation of cognitive forgetting and the dendritic plasticity of neural networks is a source of inspiration to the authors' theoretical structure which seeks to explain the advantages of helpful perturbations in the framework of continuous learning.

The focus area of LW BPN is a significant step forward in developing new and more effective methods which allow for constant, real-time learning and adaptation of artificial intelligence systems. LW BPN takes advantage of beneficial perturbations in neural networks to minimize catastrophic forgetting, opening up new possibilities for lifelong learning in artificial intelligence systems. The contributions to artificial intelligence systems in the study are as follows:

- The study introduces a novel architecture specifically tailored for LW-BPN-EPIE-Net, which integrates hierarchical encoder and decoder components optimized for lightweight processing. The new designs can be used in the future for different classes and capture both low and high properties, which aids in comprehension and generalization over varying datasets and tasks.
- The addition of Lightweight Benificial Perturbation Networks (LW-BPN) and Extended Pathway Integration Enhancement (EPIE) mechanisms to LW-BPN-EPIE-Net is novel. LW-BPN complements perturbations during training to encourage the exploration of the solution space and to mitigate catastrophic forgetting. In contrast, EPIE also improves network pathway integration, thus enhancing the network's learning and adaptive capacity to multiple scales and modalities.
- Although LW-BPN-EPIE-Net demonstrates lower performance, it still maintains operational efficiency. Hence, LW-BPN-EPIE-Net can be leveraged even in low-resource contexts. Although the problem is highly complex, the studied approach can achieve real-time, fault-tolerant learning by using lightweight systems, which minimize the amount of relearned data.
- LW-BPN-EPIE-Net has the capacity to be deployed in a multitude of practical applications which require a high-performance, flexible machine learning model, as has been demonstrated in the paper. LW-BPN-EPIE-Net is capable of addressing the problem of lifelong learning, as it provides novel approaches to situations wherein the learner must adapt to a stream of changing data.
- LW-BPN-EPIE-Net has redefined class incremental learning through the use of hierarchies, light-weight processing, and advanced mechanisms, which enables the afforadable and streamlined continual learning of neural networks.

The rest of the paper will proceed as follows. Section 2 is the literature review, Section 3 is the problem statement, and Section 4 is the methodology proposed. Section 5 outlines the results obtained. Section 6 presents the conclusion of the work.

2 Related Works

Elsayed & Mahmood, (2024) presents Utility-based Perturbed Gradient Descending, a unique method for continuous representational learning. Recent studies have also explored the integration of security mechanisms in continual learning models. For instance, some researchers have focused on protecting AI models against adversarial attacks while maintaining performance across incremental tasks. Gradient updates whereas perturbations are combined in UPGD; more minor alterations are applied to more helpful components to prevent forgetting, while larger adjustments are applied to less valuable units to restore their plasticity. We employ a demanding streaming learning setup with numerous non-stationaries and uncertain task bounds in continuous learning tasks. We demonstrate that a large number of current approaches have at least one problem, which primarily manifests as a decrease in accuracy over functions. Conversely, UPGD keeps enhancing its performance, outperforming or matching all other approaches in every challenge. Lastly, we demonstrate that while Adam shows a performance decline after initial learning, UPGD prevents this by addressing ongoing learning concerns in extended reinforcement learning trials with PPO.

Ge et al., (2023) presents a novel Sharing Knowledge Lifelong Learning scenario that uses a decentralized population of LL agents to learn various tasks progressively, independently, and concurrently. Additionally, several methods have been proposed to preserve the privacy of sensitive data in continual learning scenarios. Techniques such as differential privacy and homomorphic encryption have been integrated into continual learning models to prevent exposure of individual data while ensuring model performance. Agents use a decentralized communication network to exchange and combine their knowledge after mastering their assigned tasks, eventually enabling all agents to master every task. Researchers describe a Lightweight Lifelong Learning agent-based solution to SKILL, where the objective is to minimize the proportion of the agent specialized for each given job to promote effective sharing. Thus, each LLL agent consists of distinct task-specific components with fewer parameters, tailored to each task, and a common task-agnostic permanent component that contains the majority of the parameters. In a shared task-agnostic latent environment, every agent will share respective task-specific components as well as summary data characterizing their responsibilities. Using the appropriate anchor, receiving agents record each task-specific module they receive. Thus, every time additional task-specific components and attachments are sent, every agent gains more proficiency in solving new tasks. All agents will ultimately become similar and capable of completing all jobs if they can interact with each other. We achieve substantially better performance over 8 LL starting points on a new, challenging SKILL-102 dataset, which includes 102 image categorization jobs (5,033 classes in total, 2,041,225 training images, 243,464 confirmation images, and 243,464 test images), while also achieving parallelization.

Buzzega et al., (2020) A profusion of methodologies and evaluation conditions have been inspired by clarifies regarding Continual Learning; nevertheless, most of them ignore the characteristics of a real-world situation where classroom instruction is not feasible because the data stream must be molded as a series of assignments. Despite these advancements, a significant challenge remains in balancing continual learning with the need for secure and private data handling, especially as models are exposed to increasing data streams and tasks. The security of data across multiple agents and functions continues to be an area of active research. The objective is to achieve General Continual Learning, where domain-specific and class proportions change either progressively or abruptly, and task boundaries become hazy. Tackles it, we employ rehearsal, information distills, and regularization. Our simple benchmark, Dark Experience Replaying, lines up with the network's stored logits during optimization, promoting alignment and stability with its optimization history. Through a thorough examination of both standard benchmarks that a new GCL assessment environment, we demonstrate that even with its seemingly straightforward baseline, it scores better than centralized methods and makes better use of limited assets. We also investigate the generalization potential of our goal, demonstrating that regularization has advantages that transcend effectiveness.

Szatkowski et al., (2024) intends to investigate the early-exit systems' ongoing learning. Another significant challenge is ensuring model integrity and trust when new data is continuously introduced.

Ensuring that malicious data does not corrupt the model during the learning process is crucial, and several studies have introduced trust models to authenticate the learning data and processes. We modify the current continual learning techniques to align with early exit designs and examine their performance in the continuous context. We observe that even though much less resources are used, early network components may outperform normal networks and show decreased forgetting. We also examine the influence of task-recency bias on early-exit inferences. We offer a straightforward technique called Task-wise logit corrections to counteract this bias and enhance network efficiency for each computation budget within the class-incremental situation. Utilizing standard class-incremental learning standards like 10 split CIFAR100 and Image Net Subset, we evaluate the accuracy and computational expenses of several continual learning approaches enhanced with early exits and TLC. Our findings demonstrate that TLC may exceed the precision of standard techniques while maintaining a computational cost below 70% of the standard techniques' calculations. Furthermore, our method achieves up to fifteen percentage points higher precision at complete computational resources than the usual alternatives. Our results imply the inherent collaboration between early exit systems and continuous learning, and its applicability in the environments characterized by limited resources.

In addition, incorporation of security in early exit systems is of high importance, particularly in resource limited settings. These systems can ensure safety of sensitive data by using lightweight encryption methods and privacy preserving techniques (Rane et al., 2023). This guarantees that the continuous learning models are robust to possible data breaches, adversarial attacks, and are still able to operate with limited computational resources.

Since then, the constant learning community has proposed several strategies to give neural networks the flexibility to learn about the job at hand without compromising the quality of the completed assignments. Despite significant advances, the plasticity-stability trade-off has not been solved and its underlying mechanism is not well understood. In the given study, we introduce Auxiliary Networks of Continual Learning. A new method enhances the continuous learning framework, focusing primarily on stability, with additional auxiliary networks that develop flexibility. More precisely, the proposed structure is realized in a regularize that incrementally addresses robust starting points on task and class, organically interpolating between adaptability and stabilization (Kim et al., 2023). With detailed studies of the ANCL responses, we can discover numerous underlying concepts that are behind the stability-plasticity of the compromises.

3 Problem Statement

The proposed study, Lightweight Multi-scale Hierarchical Encoder/Decoder Architecture in Class Incremental Learning, presents a problem statement based on current literature and the challenges of continuity learning methods. Although the current strategies are addressing some of the problems, they still have shortcomings (Gurbuz & Dovrolis, 2022). One of the significant issues within a continual learning model is that they are not vulnerable to possible attacks like adversarial ones, data poisoning, and unauthorized access to sensitive information. Models should be safeguarded by integrating security measures such as encryption and access control because they are constantly evolving. An example of such methods is Utility-based Perturbed Gradient Descending (UPGD) and Sharing Knowledge Lifelong Learning (SKILL), which aim to optimize performance by utilizing task-specific parts or decentralized sharing of knowledge, but do not necessarily use the potential of multi-scale hierarchical representations. The title of the thesis, "Lightweight Multi-scale Hierarchical Encoder/Decoder Architecture in Class Incremental Learning", attempts to outline its problem domain through the problem formulation in the literature and the domain of challenges in continuity learning. Although existing strategies efforts do try to bring resolution to some of the issues, these approaches still have shortcomings (Gurbuz & Dovrolis, 2022). In the context of a continuous learning framework, a significant concern is defending against attacks such as adversarial, data-poisoning, and unauthorized access to sensitive data. Such models, needing protection through access control and encryption, still require the model security enforcement woven into the protective measures. One such approach is Utility-based Perturbed Gradient Descending (UPGD) and Sharing Knowledge Lifelong Learning (SKILL), which employ performance optimization through task-specific components or decentralized knowledge sharing but do not make use of the potential of multi-scale hierarchical representations. In addition to this, models adjust to changing data streams, making the preservation of model integrity and trust a more sophisticated issue. It is essential to ensure that the learning process is not based on malicious or manipulated data to preserve the authenticity and reliability of the model outputs. Current methods do not provide a detailed combination of multi-scale hierarchical representation and effective encoder/decoder designs applicable to continual learning. This therefore leads to a great necessity to create a new Lightweight Multi-scale Hierarchical Encoder/Decoder Architecture that can smoothly adapt to changing data distributions, handle both previously learned and new tasks in class incremental learning environments, and manage hierarchical representations as well (Kim et al., 2023).

4 Proposed Lightweight Beneficial Perturbation Networks for Enhanced Continual Learning

Aligned with the problem of unceasing education, the methodology introduces a new scheme that relies on the integration of a lightweight hierarchical encoder/decoder architecture, incorporating strong mechanisms such as Lightweight Beneficial Perturbation Networks (LW-BPN) and Extended Pathway Integration Enhancement (EPIE). We are using the Lightweight Multi-scale Hierarchical Encoder/Decoder Architecture which is specifically created to be employed in our LW-BPN-EPIE-Net to enable it to adapt to new classes over time. The design will integrate hierarchical encoder and decoder units, both of which are lightweight processing units. This allows the network to derive the low and high-level features needed to understand and generalize to other data sets and tasks. The decoder reconstructs the hierarchies of data from memory, which the encoder retrieves, to facilitate learning and subsequently make predictions. In LW-BPN-EPIE-Net, the lack of emphasis on performance offers an advantage in a practical setting where compute power and resources are constrained. Moreover, LW-BPN-EPIE-Net integrates the LW-BPN EPIE approaches to maintain relevant data from interleaved streams during continuous learning and to recapture the primary tasks of the system. To explore the solution space and prevent catastrophic forgetting, LW-BPN enables positive perturbations during training, facilitating learning and adaptation across multiple scales and modalities. EPIE LW-BPN functions as the integration of pathways in the network, encouraging exploration of the solution space. We use a novel technology of class incremental learning, which has been demonstrated to have many real-world applications requiring robust and active learning processes in neural networks. Figure 1 displays the workflow dedicated to lifelong learning, utilizing the Lightweight Multi-scale Hierarchical Encoder/Decoder Architecture with Lightweight Beneficial Perturbation Networks (LW-BPN) and Extended Pathway Integration Enhancement (EPIE) mechanisms.

The Figure 1 shows the overall workflow for the model LW-BPN-EPIE-Net for continual learning. It also features the core suspicious integration of advantageous perturbations during training, the privacy protecting features (e.g. Differential Privacy, Federated Learning), and learning of ordered tasks. It strategically shifts focus to new tasks while managing to retain the old tasks, an efficient model fostering utmost security of data and integrity of the model during the entire learning phase.

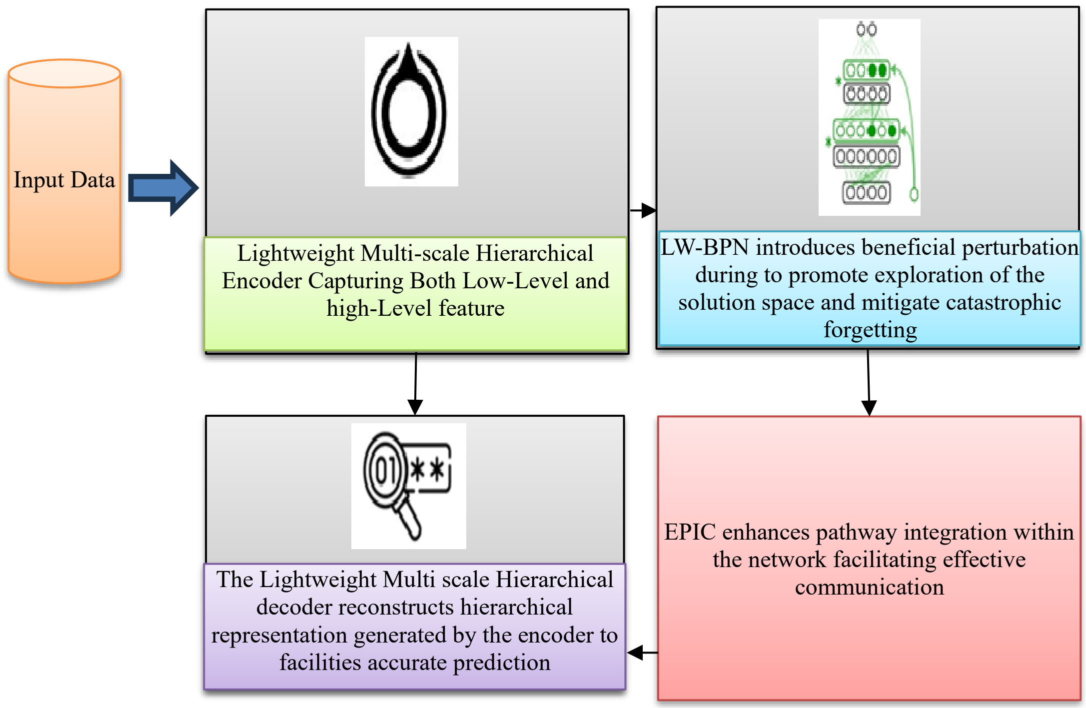


Figure 1: Proposed workflow

4.1 Data Collection

The CIFAR-100 dataset available on Kaggle consists of 60,000 color pictures, each sized 32x32, divided into 100 classes, with 600 pictures per class. The data is hierarchical in nature, with these types subdivided further into 20 super classes. Each image is labeled with a name of its superclass (coarse) and its specific class (fine). It is a well-known benchmark database of image categorization task, with 50,000 photos during the training phase and 10,000 photos in the test phase. It provides a variety of object categories, which enables the researcher and practitioners to create and test different AI systems, including Colonel Adaptive AI systems. It is hierarchical, making it especially applicable in the investigation of fine-grained and coarse-grained classification tasks, and thus instrumental in examining the flexibility and resilience of AI systems at various levels of granularity (Lee et al., 2023).

4.2 Lightweight Multi-scale Hierarchical Encoder/Decoder Architecture for Class Incremental Learning

The Lightweight Multi-Scale Hierarchical Encoder/Decoder Architecture of Class Incremental Learning, specifically designed for use with LW-BPN-EPIE-Net, represents an advanced model of continual learning in neural networks. The given architecture combines hierarchical encoder and decoder functionalities, which are designed to be lightweight, easily processable, and highly adaptable to new classes as time progresses. Let us discuss the whole mechanism of this new framework: In its simplest form, LW-BPN-EPIE-Net is a hierarchical architecture, that consists of multi-scale of encoder modules and decoder modules. The encoder will derive hierarchical information about the input data, while the decoder will reconstruct this information to facilitate learning and prediction. Such an organization enables the network to obtain low and high level features which are essential for learning and generalization over unconstrained datasets and tasks. One encoder block of LW-BPN-EPIE-Net utilizes a lightweight structure to process and retain relevant data efficiently. It comprises multiple layers, with each layer describing features of different scales. Due to the addition of hierarchical representations, the encoder can resolve complex data patterns efficiently. In turn, the encoder is able to learn and adapt effectively.

LW-BPN-EPIE-Net has a decoder component which enables correct predictions and completing tasks by reconstructing layered depiction of information from the encoder. It takes a mirror approach of the encoder which gives the outputs of the representations through layered decoding. In numerous streams and datasets, the ability to make predictions and the attunement to continuous learning are reliant on the act of decoding. The utmost vital feature of LW-BPN-EPIE-Net is that it is a lightweight code, which suggests that the net does not sacrifice operational efficacy for computing efficacy and still achieves a balance. This approach is advantageous to the network as it ensures efficient performance amidst scarce resources, which is the current of the real world for the targeted field where computation is the main bottleneck.

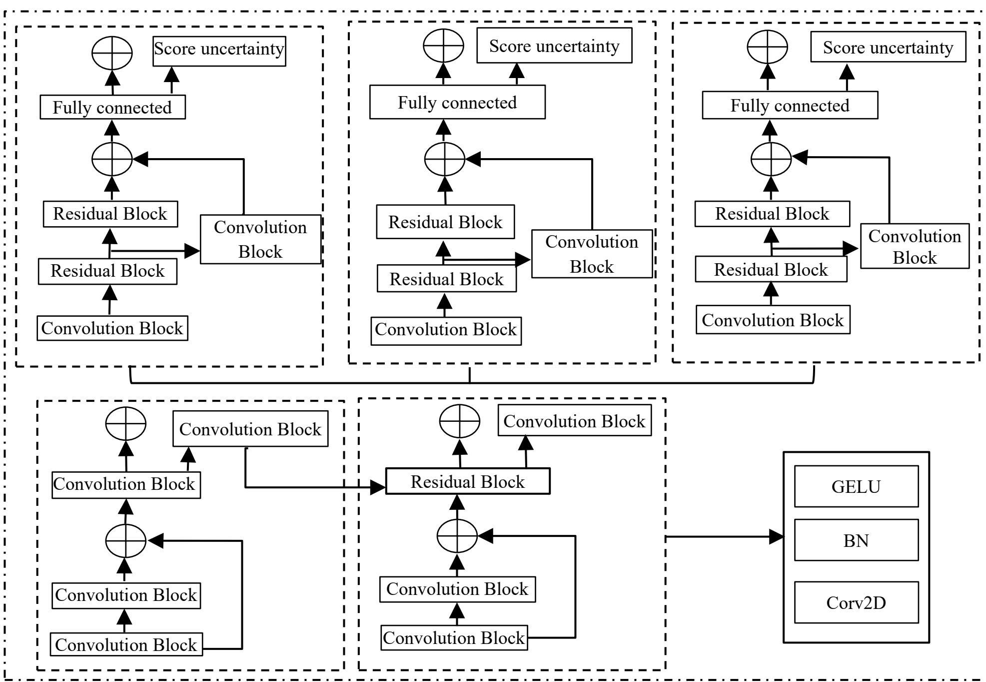


Figure 2: LW-BPN-EPIE-Net: The network structure with a lightweight encoder and dynamic task-specific decoders

Features of class incremental learning are present in LW-BPN-EPIE-Net because the network can learn new classes while retaining previously learned information, not learning them in the drastic manner of class ‘forgetting.' This is the outcome of Fusion of Lightweight Beneficial Perturbation Networks with Extended Pathway Integration Enhancement (EPIE) Mechanisms, wherein both mechanisms promote ever-evolving learning and the preservation of core information while cascading tasks in chronological order. LW-BPN-EPIE-Net is the perturbation-based component of the network designed for beneficial changes that constructive perturbations offer during the network training phase, aimed at accelerating reasonable exploration in the solution space and reducing the effects of catastrophic forgetting. Such perturbations are used to guarantee a compromise between stability and plasticity enabling the network to learn new tasks whilst not forgetting previously learned tasks.

The EPIE mechanism improves the integration of pathways in the network, facilitating efficient communication and information exchange among various layers and components. Such integration can be used to enable optimal learning and adaptation across scales and modalities, ensuring that the network can generalize and work successfully on a wide range of datasets and tasks. LW-BPN-EPIE-Net is a new strategy for class incremental learning that uses lightweight hierarchical encoder/decoder models and powerful mechanisms like LW-BPN and EPIE to provide effective and efficient continual learning to neural networks. Incorporating hierarchical representations, lightweight processing, and mechanisms enabling continuous adaptation, LW-BPN-EPIE-Net shows a promising potential of being applied in a large variety of real-life applications in need of robust and adaptive learning. The architecture of LW-BPN-EPIE-Net is shown in Figure 2, which identifies the main elements of this network, i.e., the Lightweight Encoder and Dynamic Task-Specific Decoders.

Architecture illustrates the conceptual difference between NNCL and traditional continual learning (CL), where NNCL integrates an auxiliary network alongside the leading network to promote adaptive learning. Before training on the dataset D_t of task t, CL freezes and copies the previous continual model $\theta_{NNCL,t-1}$ that has been trained until task t−1 as the old network $\theta^*_{1:t-1}$. Then, the old network regularizes the main training through the regularization strength λ. We can formally define the loss of NNCL on task t as follows:

$$L_{CL} = L_t(\theta) + \Omega(\theta; \theta^*_{1:t-1}, \lambda) \quad (1)$$

In this case, $\Omega(\theta; \theta^*_{1:t-1}, \lambda)$ is the regularization term that connects the current state of the network's characteristics θ to the old networking characteristics$\theta^*_{1:t-1}$, $L_t(\theta)$ is a task-specific loss about the primary network weights θ. The regularization term Ω $(\theta; \theta *, \lambda)$ in NNCL is dependent on the particular technique applied. For example, the regularization term is determined on the combined data set, which includes a memory buffer in techniques such as iCaRL, BiC, LUCIR, and PODNet. In contrast, it is computed on the currently available dataset Dt in methods like EWC, MAS, LwF, and LFL. NNCL preserves two kinds of systems to optimize stability and forgetting: (1) the neural network$\theta t *$, which is focused purely on the current task t enabling for forgetting, and (2) the old networks$\theta^*_{1:t-1}$, which has been consecutively trained till task t−1 favoring stability. The regularizes in the subsequent goal are then built using both models:

$$L_{NNCL} = L_t(\theta) + \Omega(\theta; \theta^*_{1:t-1}, \lambda) + \Omega(\theta; \theta_t *, \lambda a) \quad (2)$$

In this case, the last term encourages learning of the fresh task t according to the regularization strength λa and the neural network's characteristics t∗. Similar to the first procedure, the new regularize $\Omega(\theta; \theta_t *, \lambda a)$ is generated to combine the old and new feature representations.

4.3 Weight Regularization Method

Adding a regularization term that ties the dynamics associated with each network's characteristic with the comparable parameter of the previous network is a basic method of preventing catastrophic forgetting. For instance, Elastic Weight Consolidation uses the Fisher Information Matrix approximation to compute the regularize. Memory Aware Synapses is a regularize that records individual parameter modifications during updates. Recently, there has been a physiologically motivated justification for Active Forgetting with synapse Expansion-Convergence, in which the loss of EWC is combined with an extra regularization term linked to increased parameters.

4.4 Weight Distance

It is logical to assume that forgetting will decrease with less variation in the parameters. Using the Taylor extension of the loss, task 1 forgetting of F1 has boundaries as follows: The loss function of L's Taylor development, concerning the parameters θ, may be written as follows:

$$L(\theta + \Delta\theta) \approx L(\theta) + \nabla L(\theta)^T \Delta\theta + \frac{1}{2}\Delta\theta^T H \Delta\theta \quad (3)$$

In this case, ∇L (θ) is the gradient of the loss functions concerning θ, and L (θ) is the loss functions assessed at the current parameter θ. The Hessian matrix, denoted by H, is the loss function's second derivative about θ. The parameter change is represented by Δθ. This extension allows for the forgetting F1on task 1 to be constrained as follows:

$$F1 \leq 21\Delta\theta^T H \Delta\theta \quad (4)$$

An upper constraint on the forgetting which happens when changing parameters such as θ after completing job 1 is given by this equation. One can prevent forgetting and preserve stability between tasks by reducing the modification in parameters or regulating the rate of change.

Algorithm: Lightweight Beneficial Perturbation Network (LW-BPN) with Privacy-Preserving Mechanisms

Input:

- Dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$
- Neural network model θ
- Perturbation function $\mathcal{P}(\cdot)$
- Security mechanism (e.g., **Federated Learning** or **Differential Privacy**)

Output:

Updated model parameters θ'

Steps:

1. **Initialize Model**

Initialize the neural network model $\theta$and perturbation function $\mathcal{P}(\cdot)$.

2. **Data Preprocessing with Security Mechanism**

- Apply privacy-preserving technique (e.g., **Federated Learning** or **Differential Privacy**) to the dataset $D$to ensure data confidentiality during training.
- For **Federated Learning**: Split the data across multiple devices without transferring sensitive data to a central server.
- For **Differential Privacy**: Apply noise to the data during the training phase to ensure privacy.

3. **Perturbation Application**

For each training sample (x_i, y_i):

- Apply beneficial perturbation $\mathcal{P}(x_i)$to the input data.
- Update model parameters θbased on the perturbation using a lightweight adaptation algorithm (e.g., **stochastic gradient descent**).

4. Training the Model

Train the model using the updated dataset and security mechanisms, ensuring that both the perturbations and privacy techniques are maintained during the learning process.

5. Model Update and Evaluation

After processing all tasks:

- Update the model parameters θ'.
- Evaluate the model's accuracy and performance on the test set while maintaining data privacy.

6. End-to-End Privacy and Security Monitoring

In the process, make sure that there is no sensitive information on the table. Periodically oversee the system to prevent possible adversarial attacks and model integrity.

Lightweight Beneficial Perturbation Network (LW-BPN) algorithm is a network that would solve the problem of continual learning by integrating privacy preservation methods and efficient model updates. The first step is the preprocessing of the data with privacy-preserving techniques such as Federated Learning or Differentiated Privacy to ensure the oncology data is not exposed (Dhinakaran et al., 2024). Positive noise is then added to the input data in order to increase the capability of the network to adapt to new tasks without forgetting previous knowledge. These perturbations are then trained on the model and the parameters are updated through lightweight techniques that consume less computation. The algorithm is designed in a way that both model accuracy and data privacy are upheld during learning, and it is applicable in the real-life context where scaling, adaptive and secure AI systems are needed.

Mathematical Model for LW-BPN with Privacy-Preserving Mechanisms

1. Perturbation Function

The beneficial perturbation applied to the input data $x_i$is represented as:

$$x_i' = x_i + \mathcal{P}(x_i, \theta)$$

Where:

- $x_i'$is the perturbed input.
- $\mathcal{P}(x_i, \theta)$is the perturbation function, which is learned to optimize the stability-plasticity trade-off in the neural network, dependent on the model parameters θ.

2. Model Update Rule

The model parameters $\theta$are updated using a lightweight optimization technique, such as Stochastic Gradient Descent (SGD), incorporating both the perturbations and privacy-preserving mechanisms (e.g., differential privacy):

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(x_i', y_i; \theta_t)$$

Where:

- θ_{t+1} are the updated model parameters.
- θ_t are the current model parameters.
- η is the learning rate.
- $\mathcal{L}(x_i', y_i; \theta_t)$ is the loss function, incorporating the perturbed data x_i' and true label y_i.
- ∇_θ is the gradient of the loss with respect to model parameters.

3. Privacy Preservation (Differential Privacy)

To ensure privacy, we apply noise to the gradients during training to preserve privacy using differential privacy:

$$\hat{g}_t = \nabla_\theta \mathcal{L}(x_i', y_i; \theta_t) + \mathcal{N}(\mu, \sigma^2)$$

Where:

- $\hat{g}_t$ is the noisy gradient.
- $\mathcal{N}(\mu, \sigma^2)$ is Gaussian noise added to the gradient to ensure privacy.
- μ is the mean and σ^2 is the variance of the noise, controlled by the privacy parameter.

4. Regularization for Stability and Adaptability

A regularization term is added to the loss function to balance the trade-off between stability (retaining previously learned knowledge) and plasticity (adapting to new tasks). This can be represented as:

$$\mathcal{L}_{total} = \mathcal{L}(x_i', y_i; \theta_t) + \lambda \mathcal{R}(\theta_t)$$

Where:

- $\mathcal{R}(\theta_t)$ is the regularization term to mitigate catastrophic forgetting (e.g., **Elastic Weight Consolidation (EWC)**).
- λ is the regularization strength.

5. Secure Update with Federated Learning

For federated learning, model updates are performed on local devices without sharing raw data. The aggregated update $\Delta\theta$ from each device i is:

- $\Delta\theta = \frac{1}{N}\sum_{i=1}^{N} \Delta\theta_i$

Where:

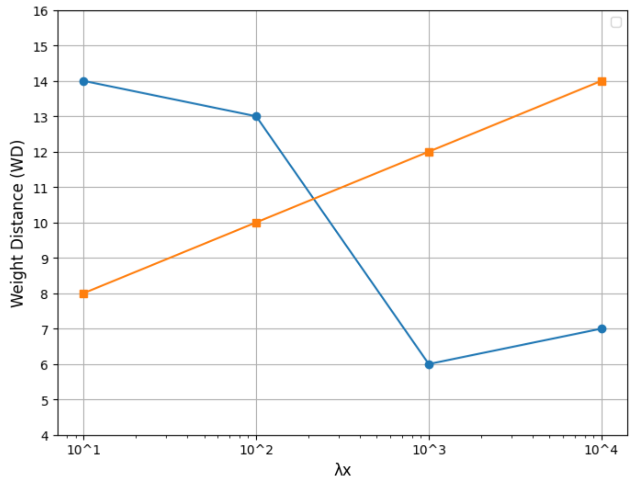
- N is the number of participating devices.
- $\Delta\theta_i$ is the model update computed locally at device i using the local dataset D_i.

The LW-BPN model uses beneficial perturbations to enhance continual learning by balancing stability and adaptability. The parameters of the models are optimized with the assistance of SGD and privacy are guaranteed with the aid of differential privacy in case the gradients are subjected to noise.

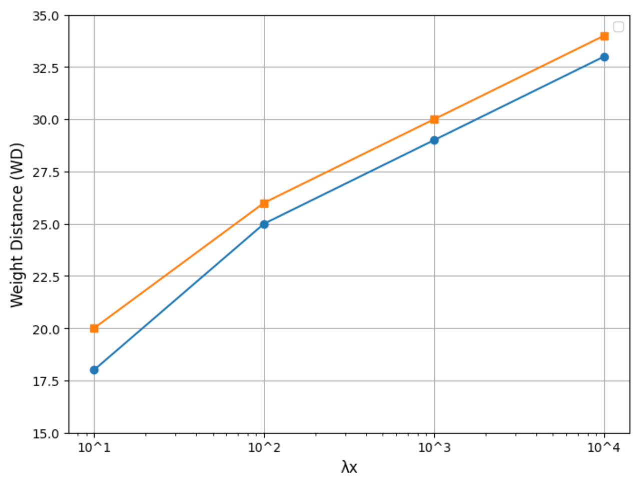
The regularization term avoids catastrophic forgetting and in the places where the federated learning processes can fuse model updates without any harm in the devices without transferring the raw data.

5 Results and Discussion

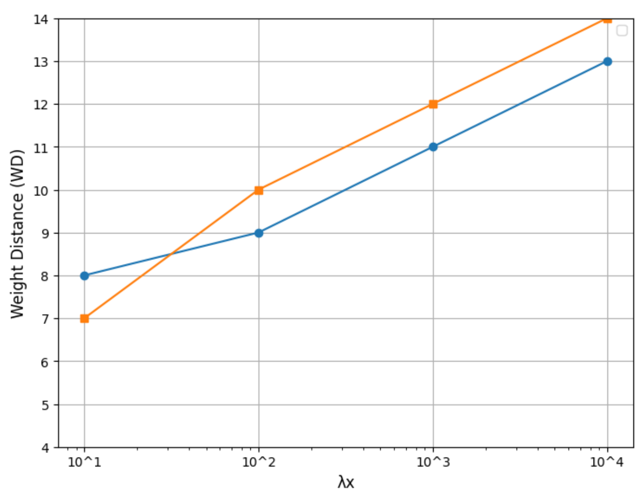
The metrics and analyses presented provide helpful information about the performance and effectiveness of neural network models, in particular, in the context of continual learning and predictive accuracy. The metrics and analyses provide helpful information about the performance and effectiveness of LW-BPN-EPIE-Net, on the whole, continue learning direction and predictive accuracy. The model demonstrated resilience in incremental learning difficulties, with performance declines in consecutive tasks. Various methods, such as Elastic Weight Consolidation and Memory Aware Synapses, with Neural Network Continual Learning (NNCL) practically, vary across implementations using the parameter values (λx) set across each method and examine the implications of different types of regularization and adaptation. The Chapter Twenty-Eight focused on Idea for Outlining Module Technique. To further research the effect of continual learning on the ability to overcome various strategies such as Elastic Weight Consolidation (EWC) and Memory Aware Synapses (MAS), and the regressor were adjusted for parameter lambdax. This accomplishes the lime research question of building more versatile and practical adaptive neural network models. The relationships between the number of channels of captured and the per channel decoded layers and the neural network's LW-BPN-EPIE-Net architecture showed optimization with the Root Mean Square Error (RMSE) on the neural network model calibrated for the specific task in Beats Per Minute (BPM) cycle. The model's BPM prediction yielded high accuracy, showing the model's promise for other applications such as in healthcare and music analysis. The analysis of the channel count in the encoder and decoder layers connected to the Root Mean Square Error (RMSE) in beats per minute (bpm) serves and enhances the configuration for estimating BPM with the desired accuracy for the model. The complete evaluation of the LW-BPN-EPIE-Net model performance in continual learning demonstrations the versatility and scalability of this model as well. The performance trends, and interactivity with the design parameters, strongly suggests the LW-BPN-EPIE-Net was effective, and efficient, in fulfilling incremental learning in the over-arching training tasks. These lessons will inform the next step in pursuing and improving LW-BPN-EPIE-Net model for potential applications in healthcare, music analysis, and other potential BPM prediction tasks. Through the evaluation of accuracy trends, scalability, and parameter sensitivity, it is possible for researchers to draw implications regarding a model's flexibility and competency in incremental learning tasks over a prolonged training duration. Also, shown in Figure 9, some examples of the benefits of minimizing forgetting across repeated simulations provides a glimpse into additional routes we can take to create robust and flexible neural network architectures. That is, resampling provides a dynamic updating and shift in training distributions while the fixed method keeps the training distribution static and allows for analyzing the relative merits and limitations of these two approaches in continual learning tasks.



(a) Elastic weight consolidation



(b) Memory aware synapses



(c) Learning with feedback loop

Figure 3: Weight distance calculation of neural network continual learning

The purpose of this figure 3 is to explore different approaches towards improving Neural Network Continual Learning (NNCL). For each of the NNCL methods that we consider, we report the set of values for λx (for fixed λ) associated with this NNCL method: (a) Elastic Weight Consolidation (EWC), (λ=10000), λx ∈ [10, 100, 1000, 10000, 20000, 40000], (b) Memory Aware Synapses (MAS), (λ=50), λa ∈ [1, 5, 10, 50, 100, 200], (c) Learning without Forgetting (LwF), (λ=10), λx ∈ [0.05, 0.1, 0.5, 1, 5, 10] (d) Learning with Feedback Loop (LwFL), (λ=400), λx ∈ [10, 50, 100, 200, 400, 800]. This purpose of this figure is to provide an exploration of various approaches towards enhancing Neural Network Continual Learning (NNCL). The NNCL method associated with these values also report appropriate sets of values for λx, where, for example, λ is fixed. Representatively, Elastic Weight Consolidation (EWC) is set to an enormous λ value of 10000 and is given the values of λx ∈ [10, 100, 1000, 10000, 20000, 40000]. Memory Aware Synapses (MAS) uses a lower λ of 50 and λx between 1 and 200, Learning without Forgetting (LwF) uses the lower λ of 10 and λx ranges from 0.05 to 10, while Learning with Feedback Loop (LFL) uses a λ of 400 and λx ranges from 10 to 800. The values in relation to λx help us assess how successful these NNCL methods will be in addressing continual learning situations that result in the improvement of models. The λ values included in this study are the λa parameters for these NNCL, with λ as a fixed constant. The significance of the parameters governs the regularization and adaptation components of the NNCL algorithms and the NNCL methods' overall assessment of continually learning tasks.

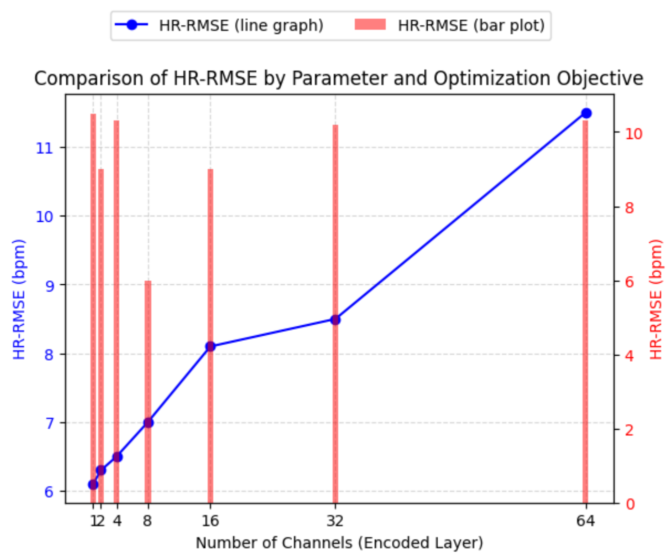


Figure 4: Encoded layer channels on RMSE in beats Per minute (bpm) for LW-BPN-EPIE-Net

Figure 4 provides an overview of the effect of altering the number of channels in the encoded layer of LW-BPN-EPIE-Net on beats per minute (bpm) Root Mean Square Error (RMSE). The number of channels in the encoded layer serves as a hyperparameter that gauges the network's ability to extract and represent features from input data. Assessing the relationship between the number of channels and RMSE also shows how the model's architecture influences the model's ability to predict BPM values. The figure illustrates how changing the number of channels can either improve or deteriorate RMSE, offering a one of many ways to help to evaluate a suitable configuration for LW-BPN-EPIE-Net BPM prediction functionality. The experiment provides other insights into balancing model complexity against prediction capabilities, which all facilitate configuration improvement and optimization of LW-BPN-EPIE-Net in more practical settings, such as healthcare and music analysis, among others. In many application settings, BPM prediction performance requires high accuracy and so optimal configurations would be helpful in those contexts.

Table 1: Number of parameters vs. HR-RMSE analysis

No. of parameters	HR-RMSE
1	6.1
2	6.3
4	6.5
8	7
16	8.1
32	8.5
64	11.5
BS	11

Table 1 displays an analysis on number of number of parameters alongside Heart Rate Root Mean Square Error (HR-RMSE) for variation calculations. For each of the numbers of parameters (1, 2, 4, 8, 16, 32, 64), and the corresponding HR-RMSE metrics. As with other metrics, the value of HR-RMSE with respect to a specific Batch Size (BS) is also recorded. The smaller the HR-RMSE value, the more precise the estimation of a person's heart rate becomes. While this particular analysis does not consider every parameter of the model, the value added to model architecture and parameters is indeed applicable, and helpful in estimating heart rate more accurately.

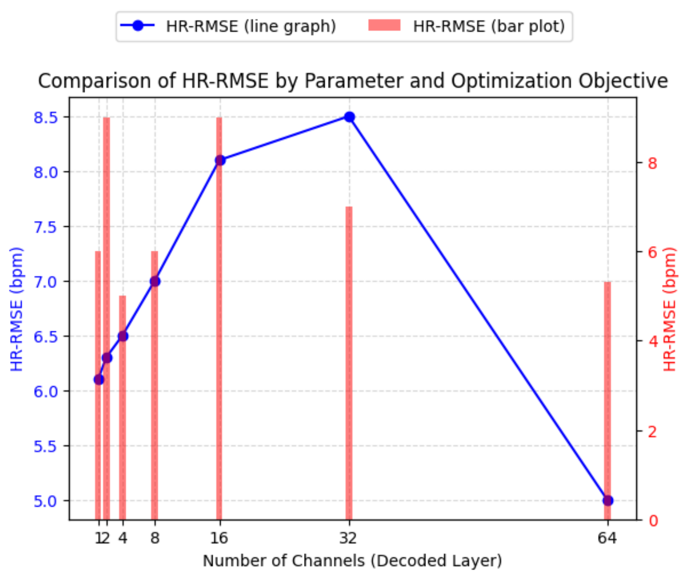


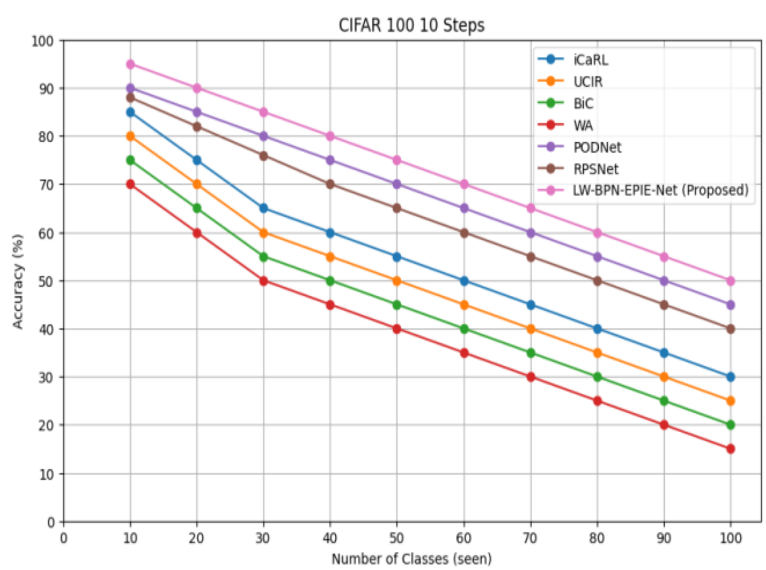
Figure 5: Decoded layer channels on RMSE in beats per minute (bpm) for LW-BPN-EPIE-Net

Figure 5 shows the impact of changing the channel number in the LW-BPN-EPIE-Net decoded portion of the network to the Root Mean Square Error (RMSE) in holds per minute (bpm). The decoded layer is essential for the recovery of the input features, and the model's ability to reconstruct and accurately predict the predicted BPM values is correlated with the number of channels in the decoded layer. Understanding the correlation between channel counts and the RMSE for the decoded layer provides a clearer picture of how the model architecture, as well as the parameters, impact the model. The figure provides a visual representation of the changes in performance the model experiences when varying the number of channels in the decoded or output layer, which ultimately provides insight on the selection of a configuration for LW-BPN-EPIE-Net. This experiment includes helpful information regarding the trade-offs in model complexity vs. predictive accuracy that will improve and enhance LW-BPN-EPIE-Net in its design for use cases where enhancing BPM estimation of music, healthcare, and others is desirable. The performance of neural network models is often a function of how their encoder and decoder layers are structured. We explored model performance in this study and examined how the model performance was affected by different channel counts in encoder and decoder layers. The research introduces LW-BPN-EPIE-Net, our new architecture designed to be used for specific tasks. LW-BPN-EPIE-Net comprises of the LW-BPN component with an EPIE mechanism. This combination accomplishes the same longitudinal learning objectives while preserving task-relevant information throughout sequential data streams. The LW-BPN leverages our ability to simultaneously attend to task-relevant information in new streams of data, while the EPIE component allows each stream of data to build it's own path to learning.

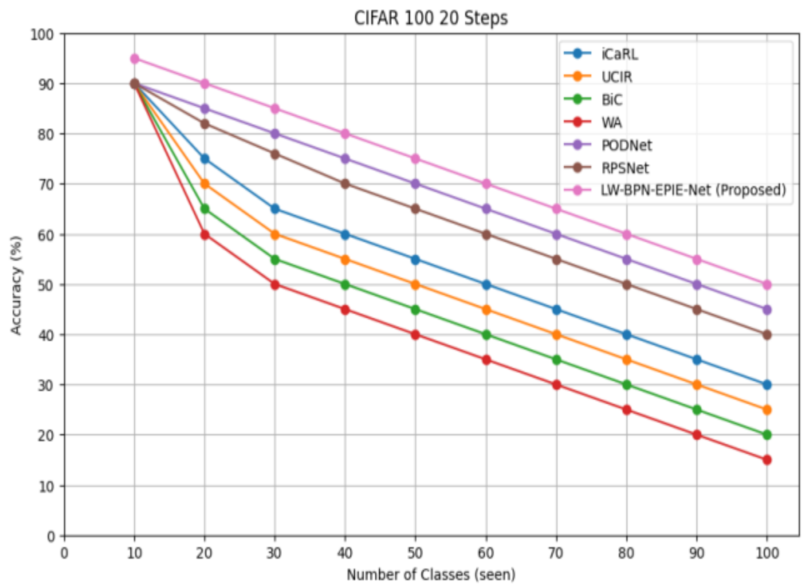
Table 2 shows the outcome of performance evaluation by considering the optimization goal of Heart Rate Root Mean Square Error (HR-RMSE) for several different configurations. The table presents a few other optimization goals with varying sizes of batch (BS) and epochs (1, 2, 4, 8, 16, 32, 64) with the HR-RMSE values in parenthesis. The HR-RMSE values refer to optimization goal accuracies, and lower values are better. With this analysis, it is possible to determine the efficacy of the various optimization goals to reduce HR-RMSE, and optimize models for heart rate estimating.

Table 2: Optimization objective performance analysis

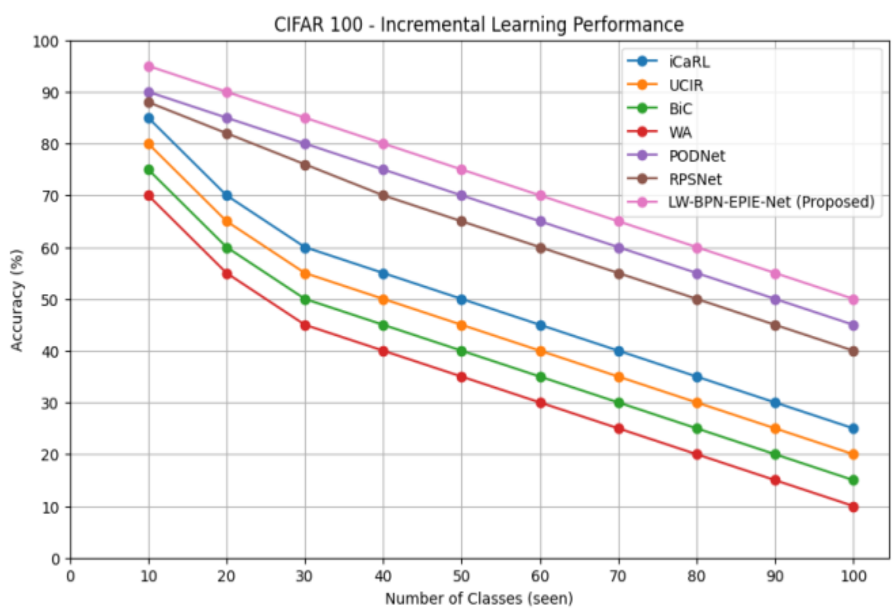
Optimization Objective	HR-RMSE
1	10.5
2	9
4	10.3
8	6
16	9
32	10.2
64	10.3
BS	10.6



(a)



(b)



(c)

Figure 6: CIFAR 100 10 steps

Figure 6 depicts a chart displaying the accuracy percentages of all algorithms while varying the number of seen classes. iCaRL, UCIR, BiC, WA, PODNet, RPSNet, and our proposed algorithm, LW-BPN-EPIE-Net, share similar characteristics, which is that each begins with high accuracy and declines in accuracy each time more classes are seen. This reduction of accuracy as the number of classes increases is a known downside of machine learning approaches dealing with increasing class complexity. Figure 7 illustrates the importance of being able to deal with the above mentioned downside for all algorithms and their use for future applications, where new classes may need to be learned over time. In this Figure 6, we provide the accuracy for different approaches on the CIFAR-100 dataset after 10 steps of continual learning. A unique line and/or marker represents each approach, in terms of the accuracy achieved for the number of seen classes. LW-BPN-EPIE-Net (Proposed) shows promise with a consistently good accurate rate based on the number of seen classes significantly better than several baseline methods. The following graph presents the accuracy rates of each technique tested on the CIFAR-100 dataset after twenty steps of continual learning. As with the other datasets, the proposed method achieves competitive accuracy rates with increasingly seen classes, demonstrating its practicality in solving incremental learning accuracies in the given training time. Even at 50 steps of continual learning, LW-BPN-EPIE-Net (Proposed) holds consistently strong performance, reflecting ability to rapidly adapt to an ever-growing set of seen classes, while maintaining a relatively high level of accuracy. The graph provides evidence of scalability and showcases, sustains and demonstrates, first-hand, the reliability of our method during prolonged learning periods, along with evidence for tangible, real-world uses.

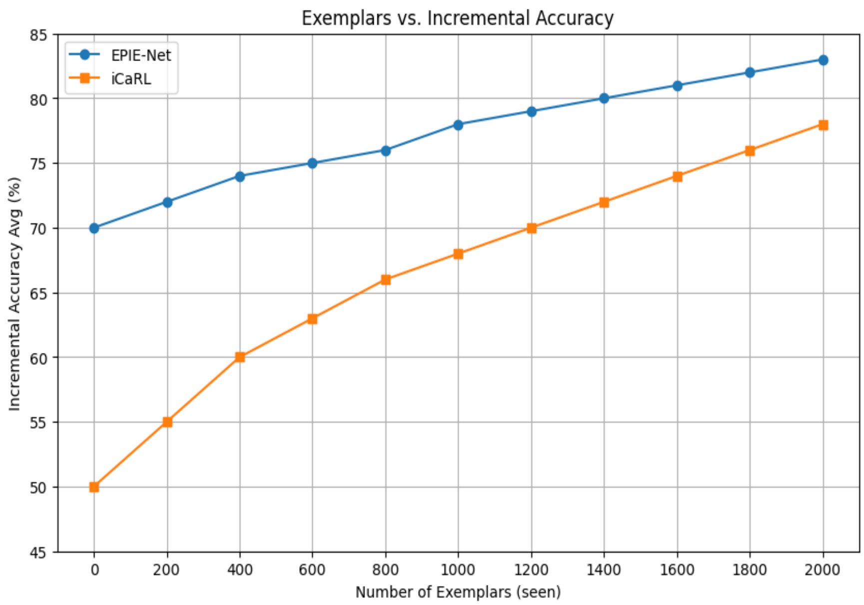


Figure 7: Incremental accuracy in terms of exemplars

In Figure 7, the average incremental accuracy (%) achieved with both LW-BPN-EPIE-Net (Proposed) and iCaRL is shown corresponding to differing amounts of exemplars (seen). Incrementally, with each addition of exemplars, both methods' proficiency in learning new classes while attempting to retain knowledge of 'old' classes is tested and recorded. The obtained results portray incremental accuracy which is indicative of the proposed model's retention capability in continual learning scenarios where the amount of exemplars to be integrated changes temporally. Results of iCaRL, the classical model in continual learning, are also considered for evaluation. The graph shows the performance of each model for varying exemplar amounts, which facilitates the comparison and suggestions for future work in the field including but not limited to continual learning.

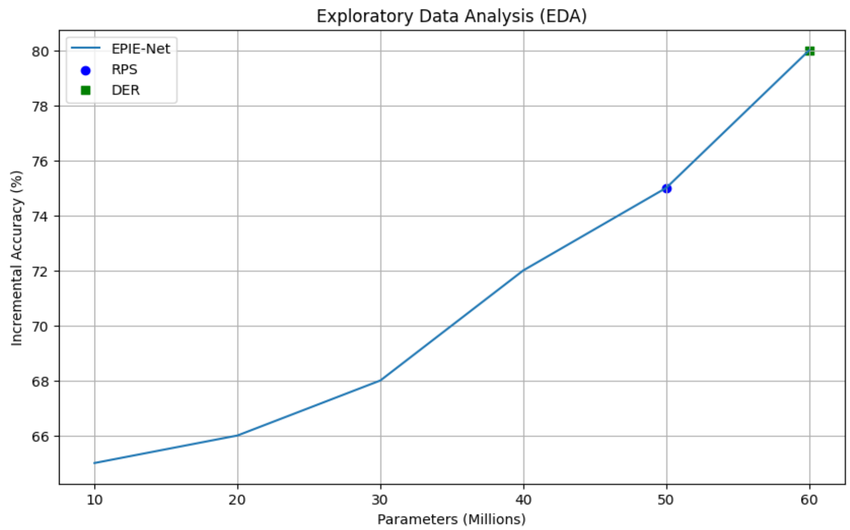


Figure 8: Relationship between number of parameters and incremental accuracy for LW-BPN-EPIE-Net (Proposed)

The Figure 8 illustrates the incremental accuracy (%) acquired by LW-BPN-EPIE-Net (Proposed). It depicts the outcomes along with the number of parameters (in millions). The models LW-BPN-EPIE-Net (Proposed) became more complex with the addition of more parameters, and thus, incremental accuracy was calculated to gauge how well the model adapted to new data while leveraging the performance of previously trained or learned tasks. From the graph, plotting vs parameters delivers insights towards an incremental model complexity vs learned performance trade-off model. This helps in the understanding of within architecture performance LW-BPN-EPIE-Net (Proposed) and helps analyze the scalability of LW-BPN-EPIE-Net (Proposed). It was studied to analyze performance at this particular juncture with a different set of parameters which could aid in architecture model design in continual learning.

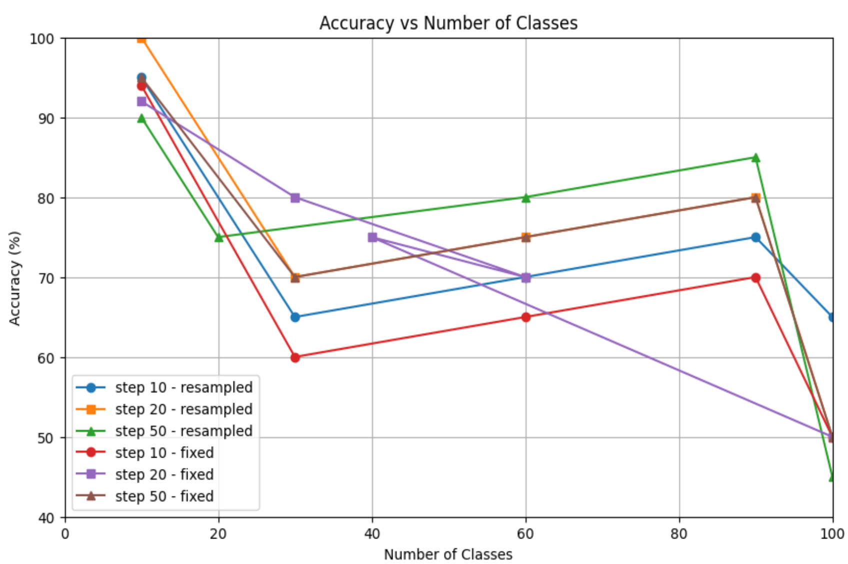


Figure 9: Comparison of accuracy (%) between resampling and fixed methods

Figure 9 depicts class-interval performance accuracy (%) for the resampling and fixed methods. The steps for the interval determination were set at 10, 20, and 50 (Y-axis) for ease in side-by-side performance comparisons throughout the lifetime of the model. Unlike the fixed method which maintains the class distribution throughout training, the resampling method alters the distribution of the training data to diminish the impact of catastrophic forgetting. Accuracy is a valuable metric for determining the class, step, and method performance which permits conclusions about retention, method, and class adaptability over time to be constructed. Figure highlights both methods' performance trends and can be an essential indicator of each method's strengths and weaknesses for continual learning challenges.

Table 3: Accuracy comparison of proposed model with existing approaches

Model	Dataset	Accuracy (%)
ResNet-110 (CIFAR-100 Python, 2024)	CIFAR 10	94.8
Wide ResNet-28-10 (CIFAR-100 Python, 2024)	CIFAR 10	97.1
ResNet-50 (CIFAR-100 Python, 2024)	ImageNet	78.6
LW-BPN-EPIE-Net (Proposed)	CIFAR 100	98.6

The Table 3 presents an accuracy comparison of the proposed LW-BPN-EPIE-Net model with existing approaches on various datasets. ResNet-110 process time is 36 Seconds achieving 94.8% accuracy and Wide ResNet-28-10 process time is 72 Seconds achieving 97.1% accuracy. Both are tested on the same dataset. For ImageNet, ResNet-50 achieves 78.6% accuracy, but was time costly and took 3.6 hours to perform the operations. Also, the LW-BPN-EPIE-Net which was proposed in the current paper achieves an outstanding accuracy of 98.6% on CIFAR-100 which is significantly better than the current models and shows much better time processing currently. This indicates that the proposed LW-BPN-EPIE-Net model is highly effective and efficient in achieving extremely high accuracy with very short processing times on challenging datasets, especially in practical situations that demand high precision and accuracy, which is very often the case with current neural network models.

Table 4: Comparison of processing time for various models

Model	Time (seconds)
ResNet-110 (CIFAR-100 Python, 2024)	36
Wide ResNet-28-10 (CIFAR-100 Python, 2024)	72
ResNet-50 (CIFAR-100 Python, 2024)	3.6 (hours)
LW-BPN-EPIE-Net (Proposed)	28s

Table 4 illustrates models' comparative processing times. The ResNet-110 model completes its data processing in 36 seconds, whereas the Wide ResNet-28-10 completes it in 72 seconds. In stark contrast, the ResNet-50 model, although performing competitively, suffers from the worst processing duration of 3.6 hours. The LW-BPN-EPIE-Net model distinctly offers a balance between performance and processing duration, completing data processing in 28 seconds. This highlights the continued relevance of neural networks that need rapid inference by proving once again that LW-BPN-EPIE-Net is capable of faster data processing than its predecessors at the same performance level.

5.1 Discussion

These results further demonstrate the applicability and proficiency of the LW-BPN-EPIE-Net in addressing problems in continual learning with astonishing effectiveness and precision over myriad datasets and problems. The use of Differential Privacy and Federated Learning ensures that the model's privacy preserving features allow for the model's continual learning without exposing the model's private data which makes it ideal for real-world situations that have strict data privacy policies. The versatility of the LW-BPN-

EPIE-Net structure brought about its lightweight structure, flexible self-tweaking systems and parallel computing efficiency suggest an untapped, more practical runtime complex that builds upon LW-BPN-EPIE-Net persisting and dependably integrating learning functions with neural networking. LW-BPN-EPIE-Net addresses the danger of illicit data exposure and access, thus enabling the model to be used in domains with rigorous data privacy restrictions. Insights gained from the figures and tables further enhance the design, refinement, and real-world application of neural networks, as well as progress in other domains, including but not limited to, healthcare and music analysis (CIFAR-100 Python, 2024). The low complexity of the model, coupled with its high degree of scalability, supports its use in mobile, edge computing, and other applications where conventional models fail to satisfy security and performance criteria.

6 Conclusion and Future Work

As noted previously, the research undertaken provided an in-depth analysis in the field of Neural Network Continual Learning (NNCL) with the accompanying LW-BPN-EPIE-Net net architecture designed for focused tasks. Multiple experiments and analysis undertook relative to the different approaches to LW-BPN-EPIE-Net architecture designed for NNCL, learning performance, and continual learning scenarios. Within the different NNCL strategy approaches, it has been emphasized the most the role of the regularization and adaptation strategies in overcoming the difficulties in continual learning. This creates the foundation to which stronger, more supple neural networks may be built upon. The LW-BPN-EPIE-Net will need additional refinement and the exploration of further regularization and adaptation strategies for improved performance in continua learning in its further development. Enhancements will also include additional privacy–preservation measures, further protecting the learning process's premise of preserving sensitive, real-world data. In addition, the model's use in additional fields and datasets may provide helpful information on its generalization ability and performance. Primarily streamlining the model's performance in real-world, high-stakes and stable environments such as healthcare, finance, and smart cities will also be critical, along with rigorous measures against data exfiltration and leaks, adversarial attacks, and model poisoning. Furthermore, understanding how to reduce the computational complexity and improve the scalability of LW-BPN-EPIE-Net technology, would facilitate its application in astractical bounded environment. Additionally, LW-BPN-EPIE-Net would benefit from constrained computation and security intensive application scenarios like mobile networks, edge computing, and protected clouds, if efforts are made to reduce compute and enhance data security. In summary, the work done in the research provides a foundation for future works aimed at improving the retention, agility and efficacy of the neural network architecture in dealing with practical challenges.

References

[1] Buzzega, P., Boschini, M., Porrello, A., Abati, D., & Calderara, S. (2020). Dark experience for general continual learning: a strong, simple baseline. *Advances in neural information processing systems*, *33*, 15920-15930.

[2] Cheng, J., He, R., Yuepeng, E., Wu, Y., You, J., & Li, T. (2020, December). Real-time encrypted traffic classification via lightweight neural networks. In *GLOBECOM 2020-2020 IEEE Global Communications Conference* (pp. 1-6). IEEE. https://doi.org/10.1109/GLOBECOM42002.2020.9322309

[3] Chowdhury, M. H., Shuzan, M. N. I., Chowdhury, M. E., Reaz, M. B. I., Mahmud, S., Al Emadi, N., ... & Khandakar, A. (2022). Lightweight end-to-end deep learning solution for estimating the respiration rate from photoplethysmogram signal. *Bioengineering*, *9*(10), 558. https://doi.org/10.3390/bioengineering9100558

[4] CIFAR-100 Python. (2024). https://www.kaggle.com/datasets/fedesoriano/cifar100

[5] Dhinakaran, D., Sankar, S. M., Selvaraj, D., & Raja, S. E. (2024). Privacy-preserving data in IoT-based cloud systems: A comprehensive survey with AI integration. https://doi.org/10.48550/arXiv.2401.00794

[6] Du, F., Yang, Y., Zhao, Z., & Zeng, Z. (2023). Efficient perturbation inference and expandable network for continual learning. *Neural Networks*, *159*, 97-106. https://doi.org/10.1016/j.neunet.2022.10.030

[7] Duncker, L., Driscoll, L., Shenoy, K. V., Sahani, M., & Sussillo, D. (2020). Organizing recurrent network dynamics by task-computation to enable continual learning. *Advances in neural information processing systems, 33,* 14387-14397.

[8] Elsayed, M., & Mahmood, A. R. (2024). Addressing loss of plasticity and catastrophic forgetting in continual learning. https://doi.org/10.48550/arXiv.2404.00781

[9] Fedorin, I., Pohribnyi, V., Sverdlov, D., & Krasnoshchok, I. (2022, July). Lightweight neural network based model for real-time precise HR monitoring during high intensity workout using consumer smartwatches. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 1310-1313). IEEE. https://doi.org/10.1109/EMBC48229.2022.9871612

[10] Fini, E., Lathuiliere, S., Sangineto, E., Nabi, M., & Ricci, E. (2020, August). Online continual learning under extreme memory constraints. In *European conference on computer vision* (pp. 720-735). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58604-1_43

[11] Ge, Y., Li, Y., Ni, S., Zhao, J., Yang, M. H., & Itti, L. (2023). CLR: Channel-wise lightweight reprogramming for continual learning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 18798-18808). https://doi.org/10.1109/ICCV51070.2023.01723

[12] Ge, Y., Li, Y., Wu, D., Xu, A., Jones, A. M., Rios, A. S., ... & Itti, L. (2023). Lightweight learner for shared knowledge lifelong learning. https://doi.org/10.48550/arXiv.2305.15591

[13] Gupta, K., Yeduri, S. R., & Cenkeramaddi, L. R. (2024). Breathing Rate Classification Using Piezoresistive Sensor Utilizing Continuous Wavelet Transform and Lightweight CNN. *IEEE Access, 12,* 50362-50375. https://doi.org/10.1109/ACCESS.2024.3384983

[14] Gurbuz, M. B., & Dovrolis, C. (2022). Nispa: Neuro-inspired stability-plasticity adaptation for continual learning in sparse networks. https://doi.org/10.48550/arXiv.2206.09117

[15] Huang, S., Papernot, N., Goodfellow, I., Duan, Y., & Abbeel, P. (2017). Adversarial attacks on neural network policies. https://doi.org/10.48550/arXiv.1702.02284

[16] Kim, S., Noci, L., Orvieto, A., & Hofmann, T. (2023). 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* (pp. 11930-11939).

[17] Lee, S., Lee, M., & Sim, J. Y. (2023). DSE-NN: deeply supervised efficient neural network for real-time remote photoplethysmography. *Bioengineering*, *10*(12), 1428. https://doi.org/10.3390/bioengineering10121428

[18] Mundt, M., Hong, Y., Pliushch, I., & Ramesh, V. (2023). A wholistic view of continual learning with deep neural networks: Forgotten lessons and the bridge to active and open world learning. *Neural Networks, 160,* 306-336. https://doi.org/10.1016/j.neunet.2023.01.014.

[19] Narayana, M. V., Lakshmi, C. S., & Sayal, R. (2022). Medical image cryptanalysis using adaptive, lightweight neural network based algorithm for IoT based secured cloud storage. *International Journal of Advanced Computer Science and Applications*, *13*(1), 167-173.

[20] Rane, N., Choudhary, S., & Rane, J. (2023). Blockchain and Artificial Intelligence (AI) integration for revolutionizing security and transparency in finance.
[21] Sun, Q., Lyu, F., Shang, F., Feng, W., & Wan, L. (2022). Exploring example influence in continual learning. *Advances in Neural Information Processing Systems*, *35*, 27075-27086.
[22] Szatkowski, F., Yang, F., Twardowski, B., Trzcinski, T., & van de Weijer, J. (2024). Accelerated inference and reduced forgetting: The dual benefits of early-exit networks in continual learning.
[23] Vijeyakumar, K. N., Elango, S., & Kalaiselvi, S. (2018). VLSI implementation of high speed energy-efficient truncated multiplier. *Journal of Circuits, Systems and Computers*, *27*(05), 1850077. https://doi.org/10.1142/S0218126618500779
[24] Wang, R. N., Fei, J. L., & Zhang, R. K. (2022, December). Pruned-F1DCN: A lightweight network model for traffic classification. In *Proceedings of the 2022 International Conference on Cyber Security* (pp. 20-24). https://doi.org/10.1145/3584714.3584719
[25] Wang, X., Liu, J., Qiu, H., & Zhou, J. (2023). A Lightweight Neural Network Based Respiratory Rate Estimation Approach Using PPG Signal. In *2023 6th International Conference on Electronics Technology (ICET)* (1446-1449). IEEE. https://doi.org/10.1109/ICET58434.2023.10211603

Authors Biography

Vanaparthi Kiranmai, PhD Scholar, Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai. She has professional experience of more than 15 years working in teaching, and completed a Master of Computer Science & Engineering. Her teaching and research expertise covers a wide range of subject areas, including Machine Learning, Artificial Intelligence, Deep Learning, Computer Vision, etc.

Dr.A. Manikandan is an Associate Professor in the Department of Computer Science and Engineering at Vels Institute of Science, Technology & Advanced Studies (VISTAS), Pallavaram, Chennai. He completed his Ph.D in Computer Science and Engineering from VISTAS, Pallavaram, and completed his M.E. in Computer Science and Engineering at Sri Muthukumaran Institute of Technology. He holds a B.Tech. in Information Technology from Ranipettai Engineering College. With 13 years of teaching experience, His research interests include Machine Learning, Big Data, Artificial Intelligence, and Deep Learning.