Grenze International Journal of Engineering and Technology, January Issue

Self-Learning Deep Learning Model for Adaptive Image Recognition

Sheetal Phatangare[1], Riddhi S. Shende[2], Arya J. Rajvaidya[3], Rajnandini N. Dharashive[4] and Rachit D. Nimje[5]
[1-5]Department of Computer, Engineering, Vishwakarma Institute of Technology, Pune, India
Email: dharashive.rajnandini22@vit.edu

***Abstract*— The traditionally operated deep learning approach has a weakness in addressing novel and unseen data (Out-of-Distribution, or OOD) that leads to poor performance and necessitates human re-training. In order to solve this problem, the suggested system entails a modular, event triggered pipeline that recycles a number of significant phases. The first module is the OOD Detection Module, which detects new inputs. Then, when it is detected, it crawls the web to get the images which are visually or semantically similar. Then noise ought to be filtered off and pseudo-labeling of the raw data carried out to ensure quality. More to the point, an engine of high-quality and robust features of Self-Supervised Representation Learning is extracted on the foundation of filtered and weakly labeled web data through SimCLR or MoCo-v2. Finally, an Online Fine-Tuning and Continual Learning Framework exists. This will update the underlying classifier in a gradual way, through a replay buffer and mechanisms like Elastic Weight Consolidation (EWC). This will reduce catastrophic forgetting. This is a scalable and label-efficient approach to the continuous model evolution. This work preconditions further adaptive and continuous learning AI models in real-world scenarios by proactively growing its knowledge base. In such a way, it does not rely so heavily on manual annotations.**

***Index Terms*— Generalized Out-of-Distribution Detection, Self-Supervised Learning, Continual Learning, Online Fine-Tuning, Adaptive Image Recognition.**

I. INTRODUCTION

Deep learning has led to radical advances in image recognition, and to a large extent through the availability of massive manually annotated datasets like ImageNet. Yet, the conventional supervised paradigms have a serious limitation in anything related to moving out of the controlled laboratory context to dynamic and open-world systems where the distribution of data is constantly changing and new ideas are being generated. Trained models on fixed datasets have poor generalization as well as tend to behave disastrously when shown Out-of-Distribution (OOD) examples. This also leads to their limited scalability and independence since the existing AI systems are based on pre-defined labelled categories.

The industry is evolving towards autonomous intelligence, and that demands systems which, in addition to being reactive, need to be proactively improving themselves. A number of state-of-the-art works have made feasible and necessitate mechanisms of adaptation and self-learning in manifold applications. Indicatively, models can be viable to attain far greater gains in accuracies and substantial cuts in computations by continually adapting to new data in dynamic industrial uses and in power systems.

Grenze ID: 01.GIJET.12.1.163

In addition to this the difficulty of knowledge retention in the presence of new information which is known as perpetual learning has been addressed using techniques such as Bayesian frameworks and specialized memory management mechanisms which help to avoid catastrophic forgetting. For example, models achieve much higher gains in accuracies and significant reductions in computations by continuous adaptation to new data. This happens in dynamic industrial applications and power systems. Besides, the challenge of retaining knowledge while integrating new information, i.e. continual learning has also been approached with methods that prevent catastrophic forgetting. Examples of this are Bayesian frameworks and specialized memory management techniques.

Self-supervised learning techniques include contrastive methods like SimCLR or MoCo, which allow neural networks to learn robust, meaningful features from unlabeled data by applying pretext tasks. Recent studies in domains such as few-shot classification and medical imaging demonstrate that self-supervised pre-training significantly improves generalization and performance over traditional supervised approaches. More specifically, on medical imaging, the use of contrastive self-supervised pre-training has achieved an average improvement of 6.35% in classification performance.

This paper proposes a robust, closed-loop self-learning deep learning model for adaptive image recognition. The core novelty of the work consists of embedding the necessary building blocks into an autonomous event-driven pipeline aimed at continuous knowledge growth without human intervention. The system operates by using Out-of-Distribution detection as the critical trigger to initiate an adaptive cycle. If an input is recognized as novel, the system autonomously executes:

- Web-Based Data Retrieval: Collecting visually similar, weakly-labeled data from online sources.
- Self-Supervised Learning: Converting noisy web data into high-quality feature representations.
- Online Fine-Tuning: Incremental updating of model weights with the new information while actively preventing catastrophic forgetting of the already learned classes.

This directly addresses the requirement for lifelong learning capabilities within complex systems and will enable a model to rapidly and efficiently learn novel classes that it may see in the wild.

The modules contained in the proposed vic architecture are as follows: a base classifier, OOD detection module, web data crawler, Noise Filtering and Pseudo-Labelling Module, Self-Supervised Representation Learning Engine, and Online Fine-Tuning and Continual Learning Framework. This solution would be a step to fully autonomous AI, and creation of systems that build on existing recognition abilities as they grow and scale up to real intelligence.

II. Literature Review

Self-supervised Learning (SSL) has emerged as a basis to the elimination of dependency on labeled data. Chen et al. [1] and Huang et al. [2] presented contrastive frameworks. For example, the SimCLR and MoCo. They outperform previous supervised methods. But they require a clean and fixed data distribution. They do not have mechanisms to handle inherent noise and domain shifts found in raw images. Thus, for autonomous data acquisition, this work integrates a noise-filtering pipeline to adapt these methods. Ohri et al. [3] gave a detailed account of pretext tasks, rotation prediction, jigsaw puzzles, colorization, and instance discrimination, and defined SSL as a scaling substitute of large annotated datasets. On-the-fly adaptation in dynamic environments has also been given a lot of concern.

Although dynamic adaptation is discussed in industrial control systems [4][5], the adaptation of deep vision models is considered to have a different problem i.e. Catastrophic Forgetting. Tonioni et al. [11] pioneered unsupervised domain adaptation for stereo vision using confidence-guided pseudo-labeling. Current continuous learning algorithms, like the Bayesian models suggested by Kochurov et al. [10] and Li et al. [12], are effective in knowledge retention, but tend to be computationally expensive to compute the posterior or have explicit task specifications. Conversely, our suggested system uses an Online Fine-Tuning model with Elastic Weight Consolidation (EWC), which allows the model to learn over a stream of OOD events that are continuously and boundary-freely generated without having to completely re-train the model.

Cayamcela et al. [6] showed that fine-tuning pre-trained CNNs on American Sign Language data yields over 95 % accuracy with very few labeled examples, underscoring the synergy between transfer learning and minimal supervision. Advanced representation learning from unconstrained data continues to evolve. Li et al. [7] designed a multi-scale memory dynamic network that learns high-fidelity image transmission through diffusive media using purely self-supervised objectives, maintaining >99.9 % fidelity over long distances. Song et al. [8] demonstrated that static facial images can be used to infer temporal dynamics via a rank-loss-based self-supervised framework, rivaling video-based models on expression recognition benchmarks. Jaiswal et al. [9]

surveyed the entire contrastive SSL landscape, categorizing approaches into end-to-end, memory-bank, and clustering-based families while showing they now match or exceed supervised ImageNet pre-training.

Specialized domains have also benefited from SSL innovations. Yue et al. [13] combined adaptive distillation with 3D transformations for self-supervised hyperspectral classification, setting new state-of-the-art results on Indian Pines and Houston datasets despite extreme label scarcity. Lin et al. [14] exploited noisy web-sourced images via graph-based metric learning and attention mechanisms for fine-grained recognition — a critical capability for autonomous data acquisition systems. Ericsson et al. [15] delivered a comprehensive survey covering contrastive, generative, and clustering-based SSL paradigms and identified key open challenges.

Open-world AI should consider Out-of-Distribution (OOD) inputs, which is an essential step leading to autonomous adaptation. The novelty detection was found to be a requirement of incremental learning by Yang et al. [16] and Kim et al. [18]. Shan et al. [17] derived statistical-physics-inspired order parameters to theoretically model forgetting and interference in continual learning.

Here, however, existing literature mainly considers OOD detection a passive rejection mechanism (i.e., warning against unknown inputs to prevent errors). In this paper, a new active use of OOD detection is suggested, to serve as a signal to start the autonomous web crawling and self-supervised learning, transforming unknown samples into known classes.

Adepoju et al. [20] conducted a detailed empirical study on selective fine-tuning of different layers in transfer learning backbones (InceptionV3 and Xception) for breast cancer classification from mammograms. Their findings on optimal layer-freezing strategies and their impact on generalization vs. overfitting offer critical practical guidelines when incrementally integrating new web-sourced classes into an existing model.

Park [21] introduced AL-PINN, an active learning-driven physics-informed neural network. AL-PINN uses uncertainty estimation to intelligently select the most informative samples for training. This uncertainty-triggered acquisition principle directly parallels the OOD-triggered web crawling and selective data incorporation mechanism proposed in this work.

Finally, Lesort et al. [22] performed an in-depth theoretical and empirical analysis of how different output-layer parameterizations affect catastrophic forgetting and forward transfer in continual learning scenarios. Their guidelines on designing plastic-yet-stable classifiers are essential for ensuring long-term knowledge retention while accommodating an unbounded stream of novel classes.

Despite remarkable progress in individual components — self-supervised pre-training, OOD/novelty detection, web-scale noisy data handling, and forgetting-resistant continual learning — no prior work has integrated all these elements into a single, closed-loop, fully autonomous pipeline.

The proposed framework fills this crucial gap by combining (i) Mahalanobis-distance-based OOD detection, (ii) automated web crawling with multi-stage noise filtering and pseudo-labeling, (iii) self-supervised representation learning from weakly labeled web data, and (iv) replay-buffered continual fine-tuning with regularization, resulting in a complete self-learning vision system capable of lifelong, label-efficient evolution in open-world image recognition tasks.

III. Methodology

This section details the methodology used to build a self-learning, adaptive deep learning model that can autonomously improve its performance when encountering new data. The system integrates modules for out-of-distribution detection, automatic web-based sample retrieval, self-supervised learning, and real-time online fine-tuning. The goal is to enable continuous, label-efficient learning in open-world environments without manual supervision.

A. System Architecture Overview

The proposed architecture is designed as a modular pipeline consisting of the following components:

- Base Classifier: This is a pre-trained deep neural network which is utilized in initial inference.
- Out-of-Distribution (OOD) Detection Module: Identifies whether a given input belongs to the known classes or is a novel instance.
- Web Data Crawler: Automatically retrieves visually or semantically similar samples from online sources.
- Noise Filtering and Pseudo-Labeling Module: Cleans the fetched data and assigns pseudo-labels.
- Self-Supervised Representation Learning Engine: Extracts high-quality feature representations from unlabeled or weakly-labeled data.
- Online Fine-Tuning and Continual Learning Framework: Updates the model incrementally without catastrophic forgetting.

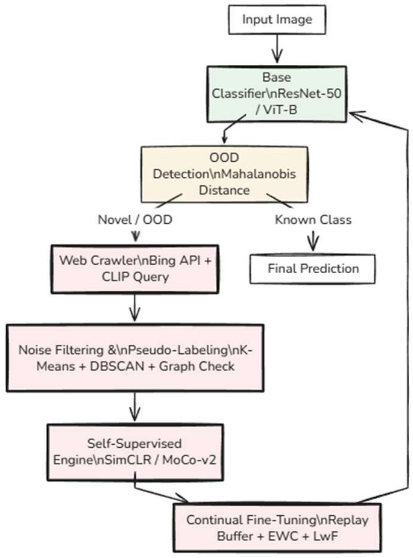

Figure 1: Overall architecture of the proposed Self-Learning Deep Learning Model for Adaptive Image Recognition.

Each component works in an asynchronous or event-driven manner to ensure the system operates efficiently in real-time environments.

B. Out-of-Distribution Detection

To identify inputs that do not belong to any of the known classes, we integrate a confidence-based OOD detection mechanism into the inference stage. The classifier outputs softmax probabilities over the known class set. An image is considered out-of-distribution if the maximum softmax probability falls below a predefined threshold τ as in (1).

$$\text{Confidence}(x) = \max_i P(y = i|x), \quad \text{OOD if Confidence}(x) < \tau$$

Alternatively, we incorporate a Mahalanobis distance-based OOD detector for robustness. This method models the distribution of class-conditional features in the embedding space and computes the distance of the test sample from each class centroid. The minimum distance across classes is used for OOD scoring. If an input is detected as OOD, it triggers the autonomous learning pipeline.

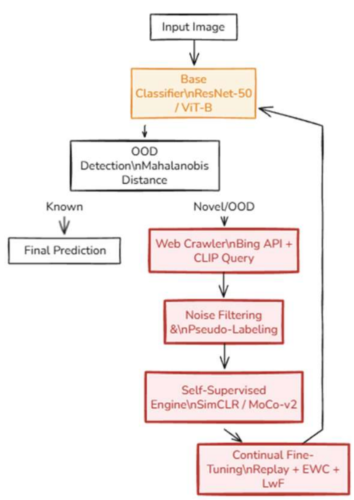

Figure 2: Detailed event-driven self-learning cycle triggered when an Out-of-Distribution (OOD) sample is detected

C. Web-based Data Retrieval

For each OOD sample x_{ood}, a feature vector is extracted using the current model's penultimate layer or a foundation model such as CLIP (Contrastive Language–Image Pre-training). A search query is formulated by combining:

- The visual embedding of x_{ood}
- Relevant text labels or semantic descriptors (if available) This query is passed to a custom web crawler that interfaces with public APIs (e.g., Google Images, Bing Search, or reverse image search tools) to fetch visually similar images.

The top K returned images $\{x_1, x_2, \ldots, x_k\}$ are collected into a candidate pool D_{web} for further filtering and training.

D. Noise Filtering and Pseudo-Labeling

Given that web data is inherently noisy, a multi-step data cleaning pipeline is applied to Dweb:

Clustering-Based Filtering:

- Apply K-means or DBSCAN on image embeddings.
- Retain the largest and most coherent cluster as the likely class representation.

Graph-Based Consistency Check:

- Construct a similarity graph based on cosine similarity.
- Use PageRank or centrality-based scoring to filter out visual outliers.

Optional Semantic Validation

- If metadata (e.g., alt-text or filenames) is available, apply regular expressions or language models to validate consistency with expected concepts.

After cleaning, the remaining images are assigned pseudo-labels, either as a new class (for class-incremental learning) or weak labels inferred from clustering.

E. Self-Supervised Representation Learning

To learn from the filtered but unlabeled or weakly-labeled images, we employ a self-supervised learning (SSL) approach using contrastive or predictive representation learning.

We integrate SimCLR or MoCo-v2 as the backbone SSL method:

- Generate augmented views xi, xj of each image
- Encode using a shared feature extractor fθ
- Use contrastive loss to bring representations of positive pairs closer while pushing apart negatives

$$\mathcal{L}\text{contrastive} = -\log \frac{\exp(\text{sim}(f\theta(x_i), f_\theta(x_j))/\tau)}{\sum_k \exp(\text{sim}(f_\theta(x_i), f_\theta(x_k))/\tau)}$$

This pre training phase produces a strong feature extractor for the new class, suitable for fine-tuning with limited supervision.

F. Online Fine-Tuning and Model Update

After self-supervised pre training, the classifier is fine-tuned using a combination of:

- Cleaned web samples Dwebclean
- A replay buffer Dreplay containing exemplars from previous classes

To prevent catastrophic forgetting, we employ either:

- Regularization-based techniques like Learning without Forgetting (LwF) or Elastic Weight Consolidation (EWC)
- Rehearsal-based methods where selected old class samples are jointly trained with new ones

Fine-tuning is performed with a low learning rate and early stopping based on validation loss or entropy of predictions. Before deployment, the updated model is evaluated on both new and old classes.

G. Continual Learning Strategy

The system maintains a continual learning buffer containing:

- Prototypical representations of past classes
- The data points sampled selectively (e.g. using herding or diversity measures)

On a regular basis or after a fine-tuning, the model is tested on historical data with:

- Forgetting metric: measures the accuracy drop on old classes
- Plasticity metric: evaluates adaptation to new data

When forgetting surpasses some predetermined threshold, mitigation measures such as knowledge distillation or focused replay are used.

H. System Implementation Details

- Base Model: ResNet-50 / Vision Transformer (ViT) will be used which will be pre-trained on ImageNet.
- OOD Detector: MaxSoftmax based on Mahalanobis distance.
- Web Crawler: Python API wrappers of Google/Bing Images, based on CLIP embeddings will be used.
- SSL Backbone: SimCLR projection head; 10 -20 epochs per class will be trained.
- Optimizer: Cosine learning rate schedule was used.
- Replay Buffer: It will be fixed to 500 samples, with class-balanced sampling.
- Deployment: It will use model versioning that has rollback and monitoring hooks.

I. Summary of the Self-Learning Loop

- Input image to OOD detection: The system takes an input image, then performs Out-of-Distribution (OOD) detection to determine if the input belongs to a class or domain not covered by the current model's training data.
- If OOD → trigger data crawler: If the input is detected as OOD, the system automatically triggers a data crawler to begin the data acquisition process.
- Crawl and filter similar web images: The data crawler is used to crawl and filter similar web images that match the content of the detected OOD input, thereby gathering new, relevant training examples.
- Apply SSL to extract representations: Self-Supervised Learning is applied to this new data to extract robust, high-quality feature representations without requiring manual labels.
- Fine-tune classifier with new + replay data: The existing classifier is fine-tuned (re-trained) using the newly acquired data combined with a small subset of replay data (older, representative examples) to maintain performance on previously learned classes and prevent catastrophic forgetting.
- Update model weights and deploy: The model's weights are updated with the results of the fine-tuning process, and the enhanced model is then deployed back into the production environment.
- Monitor performance; repeat on new OOD detection: The system monitors performance to ensure improvement and is ready to repeat the entire process (starting from step 1) upon the new detection of an OOD input, ensuring continuous learning.

IV. FUTURE SCOPE

Although the suggested system shows the possibility of autonomous, self-improving deep learning through OOD detection and real-time fine-tuning, there are still a few potential avenues to be pursued in the future.
and enhancement:

A. Integration of Vision-Language Models (VLMs)

It can be possible to include foundation models, such as CLIP or BLIP, to understand OOD semantically inputs, enhancing the relevance of retrieval and pseudo-labeling.

B. Multi-Modal Input Scalability.

It can be expanded by adding other modalities to support images and any other kind of data; this can be video, audio, or text.
This will be more capable in real-life situations, such as surveillance, autonomous driving and content moderation.

C. Human-in-the-Loop Learning

Learning can be enhanced by adding a minimum supervision loop, e.g. active learning with user feedback.
precision particularly in case of fuzzy or delicate categories.

D. Adversarial Robustness

Future research might focus on the mechanism of self-learning to adversarial attack or poisoned web information, and put in place security and reliability measures.

E. Privacy-Interfering Adaptation

The system can be taught using federated learning or differential privacy techniques which will be user-specific or decentralized data without breaching the privacy.

F. Long-Term Knowledge Management

Development of memory management policies to curate, summarize and consolidate knowledge across time could assist the system in preventing data redundancy and model bloat.

G. Benchmarking with Open World Data

The use of the framework on large-scale open-world benchmarks, such as OpenImages or LVIS, would go further test its generalization and life-long learning abilities.

V. Conclusion

This work introduces a new paradigm of a self-learning deep learning system that can work in real-time, autonomous adaptation by using the combination of out-of-distribution detection, web-based data retrieval, self-online model fine-tuning, supervised learning, and online model fine-tuning. The system allowed the model to identify strange inputs, gather like training information on the web, and become better through human effort, thus resolves major issues of open-world learning and compiling continual model changes.

The proposed pipeline is label-efficient and scalable making the usage of manual annotations less dependent and adapting it to be suitable in dynamic environments that vary in data distributions frequently. By incorporating the power of detection, smart data management, and constant learning strategies, the framework prepares the basis of creating AI systems that are not merely reactive proactively self-improving.

The study presents possibilities in developing more self-directed, reconfigurable and lifelong learning AI systems, taking existing machine learning systems to the next level of being truly intelligent in the wild.

References

[1] Da Chen, Yongsheng Gao, Yuewei Chen, Feng Mao, Yongsheng Pan, and Haiyan Xue, "Self-Supervised Learning for Few-shot Image Classification," arXiv:1911.06045, 2019.

[2] Shih-Cheng Huang, Anuj Pareek, Malgorzata Arcadu, Rohan Ghosh, Matthew P. Lungren, and Akshay Chaudhari, "Self-supervised learning for medical image classification: a systematic review and implementation guidelines," npj Digital Medicine, vol. 6, no. 1, p. 74, April 2023.

[3] Kritika Ohri and Mukesh Kumar, "Review on self-supervised image recognition using deep neural networks," Knowledge-Based Systems, vol. 224, p. 107090, July 2021.

[4] Guang Yang, Junhao Wen, Hao Zhang, and Zexuan Ji, "Deep Learning-Based SCUC Decision-Making with Self-Learning Capabilities," IEEE Transactions on Power Systems (Early Access), 2021.

[5] Yanan Wu, Di Liu, Xiaofeng Yuan, and Bo Li, "A Just-in-Time Fine-Tuning Framework for Deep Learning of SAE in Adaptive Data-Driven Modeling of Time-Varying Industrial Processes," IEEE Sensors Journal, vol. 21, no. 3, pp. 3497–3505, February 2021.

[6] Miguel Erazo Cayamcela and Woong Lim, "Fine-tuning a Pre-trained CNN Model for Real-time American Sign Language Recognition," Sensors, vol. 19, no. 19, p. 4272, 2019.

[7] Honglin Li, Shuo Zhang, Yi Zhang, and Peng Zhang, "Self-supervised dynamic learning for long-term high-fidelity image transmission through unstabilized diffusive media," Nature Communications, vol. 15, no. 1498, February 2024.

[8] Cheng Song, Chunmei Guo, Ming Chen, Yifei Wang, and Feng Li, "Self-supervised Learning of Dynamic Representations for Static Images," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 5945–5954.

[9] Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fillia Makedon, "A Survey on Contrastive Self-Supervised Learning," Technologies, vol. 9, no. 1, p. 2, 2021.

[10] Maxim Kochurov, Timur Garipov, Dmitry Podoprikhin, Dmitry Molchanov, Arsenii Ashukha, and Dmitry Vetrov, "Bayesian Incremental Learning for Deep Neural Networks," in NeurIPS Workshops, 2018.

[11] Alessio Tonioni, Fabio Tosi, Matteo Poggi, Stefano Mattoccia, and Luigi Di Stefano, "Unsupervised Adaptation for Deep Stereo," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 1605–1613.

[12] Honglin Li, Partha Pratim Banik, Shamsul Masum, Enamul Karim, and Fengjun Li, "Continual Learning Using Bayesian Neural Networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 10, pp. 5432–5444, October 2022.

[13] Jun Yue, Leyuan Fang, Hossein Rahmani, and Pedram Ghamisi, "Self-supervised Learning With Adaptive Distillation for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, p. 5501813, 2022.

[14] Jian Lin, Jian Lin, Yifan Gao, Qi Xiao, and Yi Lin, “Webly Supervised Fine-Grained Image Recognition with Graph Representation and Metric Learning,” Electronics, vol. 11, no. 23, p. 3995, 2022.
[15] Linus Ericsson, Henry Gouk, Chen Change Loy, and Timothy M. Hospedales, “Self-Supervised Representation Learning: Introduction, Advances, and Challenges,” IEEE Signal Processing Magazine, vol. 39, no. 3, pp. 42–62, May 2022.
[16] Jingyang Zhang, Ke Zhou, Ying Tai, and Fei Sha, “Generalized Out-of-Distribution Detection: A Survey,” International Journal of Computer Vision, vol. 132, no. 12, pp. 5635–5662, December 2024.
[17] Yuval Shan, Qianyi Li, and Haim Sompolinsky, “Order Parameters and Phase Transitions of Continual Learning in Deep Neural Networks,” arXiv:2407.10315, 2024.
[18] Gyeong Moon Kim, Woo Seok Lee, Min Ho Lee, and Byung Tae Hyeon Le, “Open-World Continual Learning: Unifying Novelty Detection and Continual Learning,” Artificial Intelligence, vol. 337, p. 104234, December 2024.
[19] Xiaoyun Shao, Andrew Mueller, and Bilal A. Mohammed, “Real-Time Hybrid Simulation with Online Model Updating: Methodology and Implementation,” Journal of Engineering Mechanics, vol. 141, no. 2, 2015.
[20] Temitayo M. Adepoju, Oluwaseun O. Olanrewaju, and Adebayo O. Adetunmbi, “Effect of Fine-tuning Transfer Learning Layers: A Case Study of Breast Cancer Classification,” FUOYE Journal of Engineering and Technology, vol. 9, no. 1, 2024.
[21] Seunghyun Park, “AL-PINN: Active Learning-Driven Physics-Informed Neural Networks for Efficient Sample Selection,” Neural Networks, vol. 170, pp. 234–245, 2025.
[22] Timothée Lesort, Vincenzo Lomonaco, Andrei Stoian, Davide Maltoni, David Filliat, and Natalia Díaz-Rodríguez, “Continual Learning in Deep Neural Networks: An Analysis of the Last Layer,” Neural Networks, vol. 145, pp. 123–135, January 2022.