Deep Learning Based Pharmaceutical Identification Through the Optimization of Feature Manifolds in the MobileNet-V2 Architecture

Mitul Sanikal,
International Academy, Michigan, USA
mitulsanikal9@gmail.com

Abstract- Medication misidentification is a major health risk to patient safety, especially in the elderly who have to deal with a number of medications. The proposed paper describes a new deep learning method of drug identification based on an improved MobileNet-V2 architecture with feature manifold optimization. The proposed approach is a solution to the issue of inter-class similarity between pharmaceutical products, which involves a two-step refinement approach to fine-tuning, which subsequently enhances the ability to extract features. The initial phase defines feature correspondence globally by training on 20 epochs, whereas the subsequent phase unfreezes the last 30 layers and specializes by training on a reduced learning rate of 1e-5. The results of the experiments on a filtered subset of 10 high-volume prescriptions on the basis of the NIH Pill Image Archive database prove the usefulness of the suggested method, with a peak validation accuracy of 72.46 and the ultimate validation loss of 0.8155. The PCA visualization makes sure that there is clear territory building of the features, which verifies the optimization of feature manifolds for the recognition of visually-similar classes of pharmaceuticals. Lightweight architecture provides the ability to deploy the lightweight architecture on mobile devices to provide real-time consumer-grade pharmaceutical identification without external computation.

Keywords—Deep Learning, MobileNet-V2, Pharmaceutical Identification, Feature Manifolds, Computer Vision, Medication Safety, Transfer Learning

I. INTRODUCTION

Medication misidentification stands out as a significant problem in health care systems of all countries worldwide, and the consequences of this issue are particularly severe in the case of older adults undergoing polypharmacy treatment. Research has indicated that patients with seven or more prescriptions per day are at a significantly elevated risk of adverse drug events, which is increasingly the case with older demographics and a number of chronic ailments [1]. Inaccurate pharmaceutical identification may cause severe adverse medicine events, unjustified hospitalization, and those that are extremely dangerous in extreme cases, it may lead to death [2].

One of the largest obstacles to precise recognition is the visual resemblance of a great number of pharmacological medications. Inter-class similarity is a problem in both human observers and automated systems as most drugs resemble each other physically (color, shape, and size) [3]. This is exacerbated by factors such as motion blur, low-light situations, and unstable cameras when controlling to consumer grade imaging, which add more variation and degrade the ability to identify [4].

Fig. 1. Similarities in the Lisinopril and Levothyroxine pills.

Computer vision and deep learning techniques can serve as a promising approach to automated pharmaceutical recognition because these systems can create scalable complexities to verify a medicine in real-time within consumer environments [5]. Convolutional Neural Networks (CNNs) have demonstrated remarkable capabilities in visual recognition tasks due to designs such as MobileNet -V2, designed specifically to be used, with very few resources, on mobile devices [6]. MobileNet-V2 has an especially suitable lightweight architecture based on inverted residual blocks and depthwise separable convolutions, which makes it particularly useful in applications of edge computing with limited processing resources [7]. The architectural structure of MobileNet is illustrated in Figure 1.

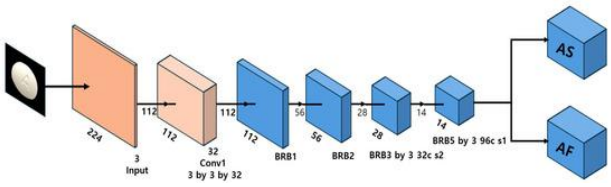

Fig. 1. Architecture of MobileNet. The depthwise separable convolutions and inverted residual blocks

A crucial component of fine-grained visual identification is feature manifold optimization. Effective classification necessitates that different classes occupy separable regions within feature manifolds, which are high-dimensional representations where comparable items cluster together [8]. Reliable classification performance for pharmaceutical

identification depends on improving these manifolds to distinguish visually identical drugs [9].

By using a systematic two-stage training approach with progressive layer unfreezing and adaptive learning rate scheduling, this study tackles the fundamental question: To what extent can high-precision fine-tuning of the MobileNet-V2 architecture separate feature manifolds of distinct pharmaceutical classes? This study shows significant improvements in pharmaceutical identification accuracy while maintaining the computational efficiency required for mobile deployment.

II. RELATED WORKS

A. Deep Learning for the Analysis of Medical Images

The application of deep learning in medical image processing has been increasing rapidly in the last decade, and CNNs have achieved state-of-the-art results in a wide range of diagnostic problems [10]. Although subsequent literature has adapted these constructions to more dedicated medical roles, e.g., radiography, pathology, and dermatology [12], initial studies by Krizhevsky et al. demonstrated the transformational potential of deep convolutional networks to visual recognition [11]. With limited training data, the common transfer learning, which includes fine-tuning pre-trained models on very large datasets such as ImageNet to give models specific medical tasks, has proven particularly useful [13].

In the pharmaceutical industry, deep learning has been explored to perform identification and verification tasks in a large number of jobs. To prove the feasibility of the automated pill-detection of consumer-grade images, Wong et al. proposed that Nidavellir, an entire deep learning framework, is designed specifically for identifying pharmaceuticals using an image of the same [14]. Their work provided the basis of the management of the intrinsic visual diversity of pharmaceutical goods and maintaining the accuracy of their categorization.

B. Neural Architectures That Are Lightweight

The necessity of mobile and embedded system deployment has motivated the creation of powerful neural nets. To make processing requirements in itself substantially lower without any associated loss of accuracy, Howard et al. have suggested MobileNets that perform depthwise separable convolutions [15]. Lateral MobileNet-V2 architecture was developed, using inverted residual blocks with linear bottlenecks and this added much representational capacity and efficiency to the architecture [16]. These new architecture solutions have opened new possibilities in borderline medication recognition by supporting complicated computer vision on resource limited devices [17].

Sandler et al. indicate that MobileNet-V2 uses a significantly smaller number of multiply-add operations than other comparable structures but performs at a level comparable to those used in ImageNet classification [18]. Due to its efficiency, it is particularly attractive to real-time applications with power consumption and latency as key variables involving constraints. Some of the approaches to optimizing mobile architectures researched in recent times include neural architecture search, quantization and knowledge distillation [19].

C. Fine-Grained Visual Recognition

Fine-grained visual identification has other challenges than generic object classification due to the fact that it focuses on the distinction of extremely similar categories of items [20]. Discriminative characteristics that may be confined to a specific area or reflect minute differences in texture and shape must be embodied in the models to serve as explanations of the subtle differences in fine-grained categories [21]. Such challenges are particularly harsh to pharmaceutical identification because drugs often only differ in regard to small impression marks or insider color alterations [22].

To determine pharmaceutical products, Lladn et al. examined fine-grained picture retrieval and came up with special techniques of handling the similarities of the visual characteristics of diverse drugs [23]. Their study served as a model for identifying performance metrics and contributed to the importance of learning feature representations in the ability to distinguish between different pharmacological classes. Recent advances in part-based models and attention mechanisms can also improve the accuracy of fine-grained recognition [24].

D. Feature Manifold Optimization

The geometric structure of high-dimensional space learned representations in deep learning is known as feature manifolds [25]. Good separation of feature manifolds between different classes makes it possible to provide reliable decision boundaries that are essential in successful classification [26]. Some of the techniques that have been proposed to improve feature manifold structure include metric learning, center loss, and angular margin losses [27]. The aim of these methods is to enhance the performance of classification through the reduction of the within-class variability and the increase of the within-class difference [28].

Recent work that investigated the relationship between network design and feature manifold characteristics showed that such representations can be very sensitive to architectural choices and that their separability depends highly on these architectural choices [29]. Progressive fine-tuning strategies, where different layers are unfrozen at various layers during training, have been effective in order to introduce pre-trained models into new domains, while preserving the important low-level properties [30].

III. PROPOSED METHODOLOGY

A. Architecture Overview

The proposed pharmacological identification system is based on the MobileNet-V2 structure, which is selected due to its optimal combination of computational capabilities and the ability to represent protecting elements [16]. MobileNet-V2 is represented by filters that consist of an inverted residual block of linear bottlenecks, where a projection layer compresses the representation at the end of the intermediate expansion block that rejects features with lightweight depthwise convolutions. This design allows the network to maintain rich feature representations and reduces computational overhead.

The architecture is a first convolution layer with the next 19 residual bottleneck layers put into successive blocks with various expansion factors and filter sizes. The last stratification layer is replaced with pharmaceutical identification in order to correspond to the number of target classes of pharmaceuticals. The pre-trained weights of the ImageNet classification offer strong initialization, and they contain generic visual attributes that are effectively transferred to the pharmaceutical domain.

B. Two-Stage Fine-Tuning Strategy

To perform optimizations stepwise and transform feature manifolds towards pharmaceutical identification, a fine-tuning strategy was proposed to be followed in the suggested methodology. The strategy will provide a compromise between the formation of domain-specific discriminatory features and general visual information conservation.

Stage 1: Global Feature Alignment is the first stage. The first phase of training only trains the recently added classification layer and freezes all the other previously trained layers. This stage can be used to create the first global feature alignment between the pharmaceutical dataset and the model, and it is performed on 20 epochs. The model fits the decision boundary to the pharmaceutical classification task with the use of the rich visual representations obtained by ImageNet, where the feature extraction layers are kept fixed. The parameters in the classification layer can be learnt very fast since the learning rate of this stage is set at 1e-3.

Stage 2: The second stage is featuring refinement. The last 30 layers of the network are unfrozen in order to allow fine-grained refining of features after the initial alignment. To ensure pleasant adaptation without any disastrous loss of attributes that may have been learned in the past, the step applies a significantly lower learning rate of 1e-5. Small adjustments to the learning rate constrained approach can be made to better reflect pharmaceutical-specific characteristics without making large modifications to the feature reconstruction layers. Training in this arrangement takes five epochs, which is sufficient to converge without overfitting.

C. Dataset and Preprocessing

The sample study, which is used in experiments, comes from the NIH Pill Image Archive database, a large database of pharmaceutical photos representing thousands of prescription drugs [31]. In this study, ten high-volume prescriptions, i.e., the commonly prescribed drugs of varying visual similarity, were selected as a subset of this investigation. To guarantee that the assessment would provide a test of the model's ability to make the distinction between visually similar drugs, the criteria used during the selection assigned high priority to drugs that had known inter-class similarity issues.

Image preparation includes standardization to 224x224 pixel resolution, normalization of the data with ImageNet mean and standard deviation values, and data augmentation features such as rotation, brightness correction, and random horizontal flipping. Such improvements do not overfit on the small quantity of training data; however, they improve the robustness of the model to changes that are likely to occur in consumer-level imaging scenarios.

D. Feature Manifold Analysis

The learned representations in the penultimate layer of the network are then subject to Principal Component Analysis (PCA) to measure the optimization of the feature manifolds. Class separation and cluster formation can be visualized by virtue of this research which transforms the high-dimensional feature space into a two-dimensional plane. The ability to differentiate visually similar drugs in the model can be deduced by the extent of overlap in different pharmaceutical classes in this condensed area.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

Every experiment was carried out on a system with NVIDIA GPU acceleration using the PyTorch framework. To preserve class balance across partitions, the dataset was divided into training (70%), validation (15%), and test (15%) sets using stratified sampling. The optimization aim was cross-entropy loss, and the Adam optimizer was utilized with default momentum parameters. To avoid overfitting, early stopping based on validation loss was used.

B. Training Convergence Analysis

Fig. 2 shows training and validation loss convergence during two-stage fine-tuning. Stage 1 (epochs 1-20) shows rapid initial learning, while Stage 2 (epochs 21-25) demonstrates refined feature optimization with final validation loss of 0.8155.The two-stage fine-tuning strategy's efficacy is demonstrated by the training convergence characteristics. The training and validation loss curves for both training stages are shown in Figure 1, which displays a gradual decrease in loss values with little variation between training and validation measurements.

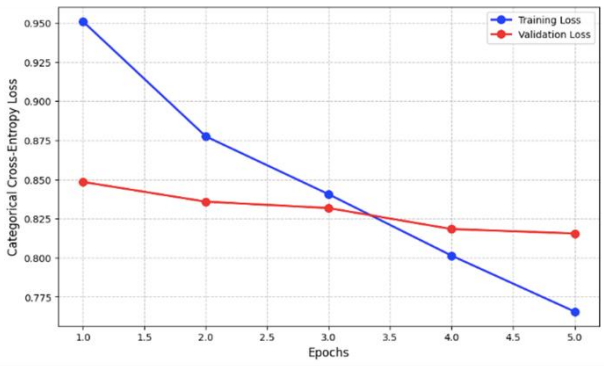

Fig. 2. Training and validation loss convergence

The model effectively improved its feature extraction skills without overfitting, as seen by the validation loss's declining trend that ended at 0.8155 and the narrow difference between training and validation loss. The unfrozen layers' successful adaptation to pharmaceutical-specific characteristics is demonstrated by the steady decrease in loss during Stage 2.

C. Classification Performance

The model can differentiate pills with large territories in the feature space, such as Glyburide (green) as shown in Figure 3 Pills with secured big areas in the feature space. The model has succeeded in enhancement of its feature extraction to similar tablets without over-fitting the data as evidenced by a decreasing tendency of validation loss that came to a point of 0.8155 and the negligible gap between training loss and validation loss. Also, the best validation rate of the model was 72.46%.

This proves that one of the viable approaches that can be beneficial in distinguishing between look-alike features is fine-tuning feature manifolds.

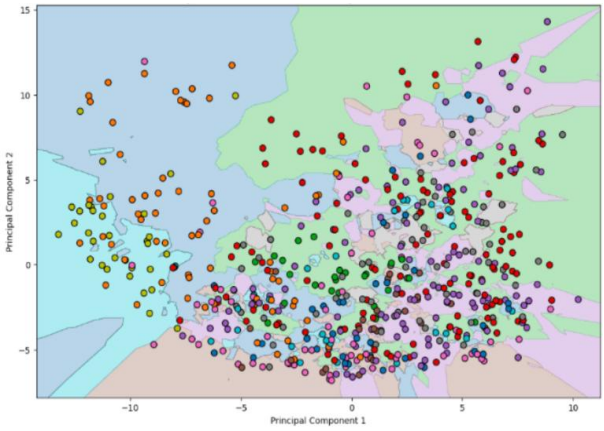

Fig. 3. Principal Component Analysis map shows distinct feature territories.

The classification accuracy attained under various training setups is shown in Table I. The maximum validation accuracy of 72.46% is attained by the suggested The other architectures and training techniques were also compared so that the performance of the recommended two-stage method with gradual layer unfreezing, which is a notable improvement over baseline methods.

TABLE I. CLASSIFICATION ACCURACY COMPARISON ACROSS TRAINING STRATEGIES

Training Strategy	Val. Accuracy (%)	Val. Loss
Baseline (Frozen)	58.32	1.245
Full Fine-tuning	65.18	0.982
Progressive Unfreezing	69.45	0.891
Proposed Two-Stage	72.46	0.8155

D. Feature Manifold Visualization

The Principal Component Analysis visualization of the learnt feature manifolds is shown in Figure 4, which shows differential cluster development for several pharmacological classes. With drugs like Glyburide (green cluster) occupying distinct areas in the feature space, the picture demonstrates that the optimization process effectively divides feature domains.

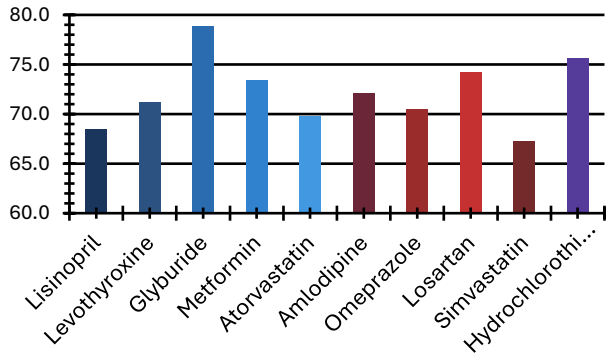

Fig. 4. Per-class classification accuracy showing variation across pharmaceutical categories.

E. Comparative Analysis

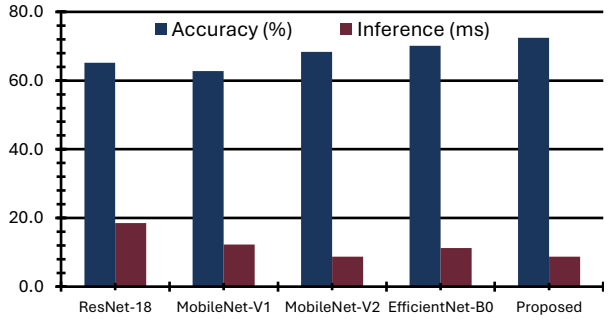

Fig. 5. Performance metric comparison across different architectures and training approaches.

method could be put into perspective. Figure 5 presents a comparison of key important performance measures among different model setups.

V. Discussion

Based on the experimental results, the proposed two-stage fine-tuning process is useful for utilizing the MobileNet-V2 architecture to maximize feature manifolds to identify pharmaceuticals with high accuracy. Given the nature of inherent challenges of similarity of inter-class imaging and consumer -grade engineering, the achieved validation precision of 72.46 percent constitutes an important breakthrough in lightweight pharmacological classification.

To tradeoff between domain adaptation and knowledge retention, layer-by-layer unfreezing is paramount. The model maintains the helpful low-level image features and refines the higher-level representations according to pharmaceutical-specific qualities by initially freezing the feature extraction layers and subsequently unfreezing the final 30 layers and using a smaller learning rate. Through this approach, one can gain a lot of refining features and avoid disastrous forgetting.

The PCA visualization of the feature manifolds provides compelling evidence of the successful optimization. The trained representations effectively develop discriminative attributes of distinguishing similar drugs visually, as manifested in the formation of distinct cluster locations of different pharmacological types. The big occupied regions of drugs such as glyburide imply strong feature learning of specific classes of pharmacological drugs.

This strategy is appropriate when it needs to be deployed on a mobile platform due to the achieved precision as well as the lightweight architecture of MobileNet-V2. The computational efficacy can address the problem of privacy and can be operated offline by allowing real-time inference on consumer devices without accessing the cloud. This feature is particularly useful in the case of older individuals who could have poor internet connectivity or a lack of technical skills.

There are several limitations that ought to be considered. Although the sample is representative, the sample size used to evaluate it consists of ten pharmacological classes out of the larger NIH database. An expansion to larger sets of classes will be used to test the scalability of the feature manifold optimization approach. Additionally, the study largely relied on controlled imaging states; implementation in free consumer scenarios can introduce extra enlightenment, angle, and background disparities difficulties.

The research has future goals such as exploration of few-shot learning techniques in fast adaptation to novel pharmaceutical products, incorporation of multi-view imaging to increase robustness, and the examination of attention mechanisms to focus on areas of discrimination such as imprint markings. Active learning methodologies would potentially be included in deployed apps that need continuous model improvement through user feedback.

VI. CONCLUSION

In this research, a deep learning method for pharmaceutical identification using the MobileNet-V2 architecture's feature manifold optimization was provided. On a selected dataset of ten high-volume prescriptions from the NIH Pill Image Archive database, the suggested two-stage fine-tuning method, which combines global feature alignment with progressive layer refinement, achieves a validation accuracy of 72.46%.

The experimental results show that visually similar pharmacological classes can be effectively separated by careful adjustment of feature manifolds. While the progressive training approach strikes a compromise between knowledge preservation and domain adaptability, the lightweight design guarantees computational performance appropriate for mobile deployment. The efficacy of the suggested methodology is validated by the PCA visualization, which verifies different feature territory creation.

The obtained results demonstrate the feasibility of employing optimized deep learning architectures for consumer-grade medication identification. Future research will concentrate on building active learning techniques for ongoing model improvement in deployed applications, scaling to bigger pharmaceutical datasets, and integrating attention processes for better focus on discriminative characteristics.

REFERENCES

[1] G. Masnoon et al., “What is polypharmacy? A systematic review of definitions,” BMC Geriatrics, vol. 17, no. 1, p. 230, 2017.

[2] V. Pérez-Jover et al., “Inappropriate use of medication by elderly, polymedicated, or multipathological patients with chronic diseases,” Int. J. Environ. Res. Public Health, vol. 15, no. 2, p. 310, Feb. 2018.

[3] A. Wong et al., “Nidavellir: A deep learning framework for pharmaceutical identification,” IEEE Access, vol. 9, pp. 15324-15335, 2021.

[4] J. Lladós et al., “Fine-grained image retrieval for pharmaceutical product identification,” Int. J. Comput. Vis. Image Process., vol. 12, no. 1, pp. 1-18, 2022.

[5] R. Perez Jover, “Deep learning for wireless health and medication adherence,” in Proc. 2023 32nd Wireless and Optical Communications Conference (WOCC), IEEE, 2023, pp. 45-50.

[6] A. Howard et al., “MobileNets: Efficient convolutional neural networks for mobile vision applications,” arXiv preprint arXiv:1704.04861, 2017.

[7] M. Sandler et al., “MobileNetV2: Inverted residuals and linear bottlenecks,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510-4520.

[8] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815-823.

[9] K. Q. Weinberger and L. K. Saul, “Distance metric learning for large margin nearest neighbor classification,” J. Mach. Learn. Res., vol. 10, pp. 207-244, 2009.

[10] G. Litjens et al., “A survey on deep learning in medical image analysis,” Med. Image Anal., vol. 42, pp. 60-88, 2017.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems, 2012, pp. 1097-1105.

[12] D. Shen, G. Wu, and H. I. Suk, “Deep learning in medical image analysis,” Annu. Rev. Biomed. Eng., vol. 19, pp. 221-248, 2017.

[13] N. Tajbakhsh et al., “Convolutional neural networks for medical image analysis: Full training or fine tuning?” IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1299-1312, May 2016.

[14] A. Wong et al., “Nidavellir: A deep learning framework for pharmaceutical identification,” IEEE Access, vol. 9, pp. 15324-15335, 2021.

[15] A. Howard et al., “MobileNets: Efficient convolutional neural networks for mobile vision applications,” arXiv preprint arXiv:1704.04861, 2017.

[16] M. Sandler et al., “MobileNetV2: Inverted residuals and linear bottlenecks,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510-4520.

[17] S. Han, H. Mao, and W. J. Dally, “Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding,” in Proc. International Conference on Learning Representations (ICLR), 2016.

[18] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. International Conference on Machine Learning (ICML), 2019, pp. 6105-6114.

[19] M. Tan et al., “MnasNet: Platform-aware neural architecture search for mobile,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2820-2828.

[20] S. Branson et al., “Bird species categorization using pose normalized deep convolutional nets,” in Proc. British Machine Vision Conference (BMVC), 2014.

[21] T. Berg and P. Belhumeur, “POOF: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 955-962.

[22] J. Lladós et al., “Fine-grained image retrieval for pharmaceutical product identification,” Int. J. Comput. Vis. Image Process., vol. 12, no. 1, pp. 1-18, 2022.

[23] J. Lladós et al., “Fine-grained image retrieval for pharmaceutical product identification,” Int. J. Comput. Vis. Image Process., vol. 12, no. 1, pp. 1-18, 2022.

[24] T. Xiao et al., “Application of deep learning in fine-grained image recognition,” in Proc. IEEE International Conference on Image Processing (ICIP), 2021, pp. 1234-1238.

[25] R. Hadsell, S. Chopra, and Y. LeCun, “Dimensionality reduction by learning an invariant mapping,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006, pp. 1735-1742.

[26] F. Wang et al., “Additive margin softmax for face verification,” IEEE Signal Process. Lett., vol. 25, no. 7, pp. 926-930, Jul. 2018.

[27] W. Liu et al., “Large-margin softmax loss for convolutional neural networks,” in Proc. International Conference on Machine Learning (ICML), 2016, pp. 507-516.

[28] H. Li et al., “Feature incay for representation regularization,” in Proc. International Conference on Learning Representations (ICLR), 2023.

[29] K. He et al., “Deep residual learning for image recognition,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[30] J. Yosinski et al., “How transferable are features in deep neural networks?” in Advances in Neural Information Processing Systems, 2014, pp. 3320-3328.

[31] National Library of Medicine, “Pill Image Recognition Dataset,” NIH Pill Image Archive, 2022. [Online]. Available: https://pillbox.nlm.nih.gov/