International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Automated Lung Cancer Classification from CT Scans using Deep Learning and Knowledge Distillation

Zuhair Ansari

Department of BECHLOR OF VOCATIONAL IN ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Anjuman-I-Islam's AbduL Razzaq Kalsekar Polytechnic, New Panvel, Maharashtra, India

---

Abstract - Lung cancer remains a leading cause of cancer-related mortality worldwide, primarily due to late-stage diagnosis. Early and accurate detection from Computed Tomography (CT) scans is critical but is often hindered by the time-consuming and subjective nature of manual interpretation. This paper presents a deep learning-based system for the automated classification of lung CT images into Normal, Benign, and Malignant categories. We leverage an MLOps-driven pipeline to ensure reproducibility and scalability. The system employs transfer learning with VGG16 and Vision Transformer (ViT) architectures and introduces knowledge distillation to train a lightweight, efficient student model (VGG16) from a high-performing teacher model (ViT). Our dataset of 2,097 CT images was balanced using data augmentation and oversampling. The final distilled VGG16 model achieves a test accuracy of 98.09%, matching the performance of the larger teacher model while being significantly more efficient. This demonstrates that knowledge distillation can produce clinically viable models that balance high accuracy with the practical requirements for deployment in real-world healthcare settings.

Key Words: Lung Cancer, Deep Learning, Computed Tomography (CT), Image Classification, VGG16, Vision Transformer (ViT), Knowledge Distillation, MLOps.

1.INTRODUCTION

Lung cancer is the most fatal cancer globally, with a starkly low survival rate that is directly linked to late-stage diagnosis [1]. Early detection is the most effective strategy to improve patient outcomes. Computed Tomography (CT) is the clinical standard for thoracic imaging, providing detailed cross-sectional images that can reveal nascent pulmonary nodules. However, the manual review of these scans is a significant bottleneck in the diagnostic pathway. Radiologists must meticulously analyze hundreds of slices per patient, a process that is not only labor-intensive but also susceptible to perceptual errors, fatigue, and high inter-observer variability [2].

To address these challenges, this project develops an automated system for classifying lung CT scan images into three clinically relevant classes: Normal, Benign, and Malignant. The primary goal is to create a reliable decision-support tool for radiologists that can accelerate diagnosis, improve consistency, and enhance accuracy.

A secondary but equally important objective is to build this system within a robust Machine Learning Operations (MLOps) framework. This ensures that the entire development lifecycle—from data ingestion and training to evaluation—is reproducible, traceable, and scalable. By using tools like DVC for pipeline orchestration and MLflow for experiment tracking, we establish a production-ready workflow. Furthermore, we explore advanced techniques like knowledge distillation to create a model that is not only accurate but also computationally efficient, making it suitable for deployment in resource-constrained clinical environments.

II. RELATED WORK

The effort to automate the analysis of medical images is not new. Early attempts relied on classical machine learning and computer-aided diagnosis (CADx) systems.

A. Classical Machine Learning Approaches Traditional CADx systems were based on a "handcrafted" feature engineering pipeline [3]. This involved complex, multi-stage processes including image preprocessing, nodule segmentation, and the manual extraction of quantitative features (e.g., size, sphericity, texture). These features were then fed into classifiers like Support Vector Machines (SVMs) or Random Forests. While pioneering, these systems were often brittle, highly dependent on the accuracy of the segmentation step, and struggled to generalize to new data from different scanners or hospitals.

B. The Rise of Deep Learning The advent of deep learning, particularly Convolutional Neural Networks (CNNs), marked a paradigm shift. Architectures like VGG16 [4], with their ability to learn hierarchical features directly from raw pixel data, demonstrated superior performance. VGG16's deep stack of small 3x3 convolutional filters proved effective for image classification and became a popular choice for transfer learning in the medical domain.

More recently, the Vision Transformer (ViT) [5], adapted from natural language processing, has emerged as a powerful alternative. By using a self-attention mechanism, ViT can model long-range dependencies across the entire image from its first layer, giving it a global receptive field that is highly advantageous for interpreting complex patterns in medical images.

C. Knowledge Distillation for Clinical Efficiency While large models like ViT achieve high accuracy, their size and computational cost can be prohibitive for clinical deployment. Knowledge distillation [6] offers a solution. In this teacher-student paradigm, the knowledge from a large, high-performing "teacher" model is transferred to a smaller, more efficient "student" model. The student is trained to mimic the teacher's output probability distribution, allowing it to learn a more effective generalization policy than it could from hard labels alone. This enables the development of models that are both accurate and practical.

III. METHODOLOGY

The system was developed using a structured, modular MLOps pipeline to ensure rigor and reproducibility. This section details the data preparation, model architectures, and training strategies employed.

A. Dataset and Preprocessing

The dataset consists of 2,097 2D CT scan image slices, categorized into Benign, Malignant, and Normal cases. The data was partitioned using stratified splitting into training (1341 images), validation (336 images), and test (420 images) sets.

The initial training set was imbalanced. To address this, we employed data augmentation and oversampling. Augmentation techniques applied to the training set included:

• Random horizontal flipping
• Random rotation (up to 15 degrees)
• Random zoom and shifts

Oversampling was then used to augment the minority classes (Malignant and Normal) to match the sample count of the majority class (Benign), resulting in a balanced training set of approximately 600 images per class. All images were resized to 224x224 pixels and normalized.

B. MLOps Pipeline

The project was architected as a Directed Acyclic Graph (DAG) of automated stages orchestrated by DVC. Key stages included data ingestion, augmentation, model training, and evaluation. MLflow was integrated to log all experiments, including hyperparameters, performance metrics, learning curves, and model artifacts. This setup provides a complete, traceable record of every experiment, facilitating data-driven model selection.

C. Model Architectures

1. VGG16 (Student Model): We used a VGG16 model pretrained on ImageNet. The convolutional base was frozen, and the final fully connected classifier was replaced with a new head adapted for our 3-class problem. This served as both a performance baseline and the "student" in our distillation experiments.

2. Vision Transformer (ViT - Teacher Model): A ViT model, pretrained on a large corpus of medical images, was selected as our high-performance "teacher" model. Its self-attention mechanism is well-suited to capturing the global context and subtle textural details present in CT scans.

D. Knowledge Distillation

To create a final model that is both accurate and efficient, we used knowledge distillation. The goal was to transfer the diagnostic capabilities of the large ViT teacher model to the lightweight VGG16 student model. The student was trained using a composite loss function:

$L_{total} = \alpha * L_{CE} + (1 - \alpha) * L_{Distill}$

where:

• L_{CE} is the standard Cross-Entropy loss between the student's predictions and the true labels.
• $L_{Distill}$ is the Kullback-Leibler (KL) Divergence loss between the softened probability distributions of the teacher and student models.
• α is a hyperparameter that balances the two loss terms.
• The softening is controlled by a temperature parameter, T.

This process guides the student to learn the nuanced decision boundaries of the teacher, resulting in improved performance.

IV. EXPERIMENTAL SETUP

All experiments were conducted using Python with the PyTorch deep learning framework. Training was performed on Google Colaboratory using NVIDIA T4 GPUs. The models were evaluated on the held-out test set using standard classification metrics: Accuracy, Precision, Recall, and F1-Score. The confusion matrix was also analyzed to understand specific error patterns.

V. RESULTS AND DISCUSSION

The performance of each model was systematically evaluated on the test set. The results are summarized in Table I.

Model	Accuracy	Precision (Malignant)	Recall (Malignant)	F1-Score (Malignant)
VGG16 (Transfer Learning)	97.14%	0.96	0.96	0.96
ViT (Teacher)	98.09%	0.97	0.98	0.98
VGG16 (Distilled Student)	98.09%	0.97	0.98	0.98

The baseline VGG16 model achieved a strong accuracy of 97.14%, demonstrating the effectiveness of transfer learning. The larger ViT teacher model improved upon this, reaching 98.09% accuracy with a notably higher recall for the critical Malignant class.

Remarkably, the distilled VGG16 student model matched the 98.09% accuracy of its much larger teacher. It inherited the teacher's ability to correctly identify malignant cases while retaining the computational efficiency of the VGG16 architecture. Analysis of the confusion matrix for the distilled model revealed only 2 false negatives for the Malignant class out of 112 actual cases, highlighting its clinical reliability.

These results validate our hypothesis that knowledge distillation is an exceptionally effective technique for developing high-performance, lightweight models for medical image analysis.

VI. FUTURE WORK

While the current system is robust, several avenues for future work exist. These include integrating Explainable AI (XAI) techniques like Grad-CAM to visualize model decisions, expanding the dataset with more diverse and rare cases, and exploring federated learning to train models across multiple institutions without sharing sensitive patient data. The ultimate goal is to deploy the system in a real-time clinical setting for prospective validation.

VII. CONCLUSION

This project successfully developed and validated an end-to-end deep learning pipeline for automated lung cancer classification from CT scans. By leveraging knowledge distillation, we produced a lightweight VGG16 model that achieves a state-of-the-art accuracy of 98.09%, matching the performance of a much larger Vision Transformer. The integration of an MLOps framework ensures that the system is reproducible, scalable, and ready for clinical deployment. This work serves as a strong proof-of-concept for how advanced AI techniques can be

practically applied to create efficient and reliable tools to aid radiologists, accelerate diagnosis, and ultimately improve patient outcomes in the fight against lung cancer.

REFERENCES

[1] World Health Organization, "Cancer," WHO Fact Sheet, 2022. [Online].
[2] J. G. Elmore et al., "Variability in Interpretive Performance at Screening Mammography and Radiologists' Characteristics Associated with Accuracy," Radiology, vol. 253, no. 3, pp. 641–651, 2009.
[3] G. D. Rubin, "Computer-Aided Detection and Diagnosis of Pulmonary Nodules," Radiologic Clinics of North America, vol. 56, no. 4, pp. 581–598, 2018.
[4] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in International Conference on Learning Representations (ICLR), 2015.
[5] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in International Conference on Learning Representations (ICLR), 2021.
[6] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," in NIPS Deep Learning and Representation Learning Workshop, 2015.