IEC
Institute of Electronics and Computer

Journal of Artificial Intelligence and Systems, 2026, 8, 95-105
https://iecscience.org/journals/AIS
ISSN Online: 2642-2859

EfficientNetB0 for Lung Cancer Classification

Razafimanana Geraldo[1,*], Rasoamanana Marie Anna Chrishna[2], Rahajaniaina Andriamasinoro[3]

Faculty of Sciences and Technology, Department of Mathematics, Computer Sciences and Applications, University of Toamasina, Toamasina, Madagascar
[1]E-mail: raharimananageraldo@gmail.com
[2]E-mail: marieannarasoamanana@gmail.com
[3]Email: hajatoam@gmail.com
*Corresponding author: Razafimanana Geraldo, raharimananageraldo@gmail.com

How to cite this paper: Razafimanana Geraldo, Rasoamanana Marie Anna Chrishna, Rahajaniaina Andriamasinoro (2026). EfficientNetB0 for Lung Cancer Classification. Journal of Artificial Intelligence and Systems, 8, 95–105. https://doi.org/10.33969/AIS.2026.080106.

Received: August 5, 2026
Accepted: November 18, 2026
Published: November 23, 2026

Copyright © 2026 by author(s) and Institute of Electronics and Computer. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

Lung cancer causes more deaths every year than any other cancer, and getting an accurate, automated read on histopathology images remains a pressing clinical need. In this work we build a deep learning pipeline that sorts lung tissue images into six categories — BenignCases, MalignantCases, adenocarcinoma, large cell carcinoma, squamous cell carcinoma and normal tissue. EfficientNetB0 serves as the backbone and is trained with an ArcFace metric-learning loss to produce 512-dimensional embeddings, discriminative enough to be clustered afterwards with K-Means and HDBSCAN. We validate the approach with 10-fold stratified cross-validation on 3,174 lung histopathology images taken from a public Kaggle dataset. The supervised model reaches 98.10% test accuracy, and both clustering algorithms independently recover the six classes with roughly 99.8% accuracy. UMAP projections of the embedding space show six compact, well-separated clusters, and HDBSCAN finds all six groups on its own, with zero noise points in any fold. Overall, these results suggest that pairing metric learning with density-based clustering is a workable route toward accurate, label-efficient lung cancer diagnosis.

Keywords

Lung cancer classification, EfficientNetB0, ArcFace, metric learning, HDBSCAN, K-Means, deep embeddings, 10-fold cross-validation, medical image analysis.

1. Introduction

Lung cancer kills more people every year than any other cancer, accounting for roughly 18% of all cancer-related deaths worldwide. Getting the pathological classification right early on matters a great deal, since it shapes treatment choices and, ultimately, survival outcomes [1]. Histopathology remains the diagnostic gold

standard, but reading these slides is a subjective and time-consuming task that depends on specialist expertise — expertise that is far from evenly distributed across health systems, particularly in low-resource settings such as ours in Madagascar.

Deep learning has changed computational pathology considerably over the past decade. Its core principles — learned hierarchical representations trained end-to-end on raw pixel data [2] — have proven remarkably transferable from natural images to clinical ones. Convolutional neural networks (CNNs) now reach strong performance on medical image classification and can process large volumes of images automatically. Most of the best-performing systems, however, are trained on large labeled datasets and rely on architectures that are computationally heavy, which is not always practical in constrained clinical settings. Labeling pathology images is itself expensive, since it requires trained specialists — one of the reasons semi-supervised and unsupervised approaches have become an active line of research.

We address both problems with a three-stage pipeline: first, supervised metric learning using EfficientNetB0 with an ArcFace loss head; second, extraction of the resulting embeddings; and third, unsupervised clustering with K-Means and HDBSCAN, all evaluated through 10-fold stratified cross-validation. ArcFace was originally designed for face recognition, where it enforces a large angular margin between classes in the embedding space. That property turns out to be useful here as well: it yields embeddings that are discriminative enough to be separated by unsupervised clustering, without the clustering algorithm ever seeing a label.

Concretely, this work contributes the following:

1). A complete, reproducible pipeline for six-class lung cancer classification that combines EfficientNetB0 with ArcFace metric learning;

2). Evidence that ArcFace embeddings are discriminative enough to support near-perfect unsupervised clustering at inference time, without any label information;

3). A rigorous, statistically grounded comparison of K-Means and HDBSCAN under 10-fold stratified cross-validation;

4). An analysis of how class imbalance affects clustering quality, together with practical pre-processing recommendations.

The rest of the paper is organized as follows. Section 2 reviews related work on CNN-based medical image classification, metric learning and unsupervised clustering. Section 3 details the proposed methodology. Section 4 describes the experimental protocol. Section 5 presents and discusses the results for both the balanced and imbalanced data configurations. Section 6 discusses the findings and their limitations, and Section 7 concludes the paper.

2. Related Work

2.1. Deep Learning for Lung Cancer Diagnosis

CNNs are now a standard tool for lung histopathology and CT-based image classification. Architectures such as VGG, ResNet [3] and DenseNet have all shown solid results on multi-class cancer classification tasks. Because labeled medical data is scarce, many recent studies rely on transfer learning from ImageNet-pretrained weights rather than training from scratch. EfficientNet, introduced by Tan and Le [4],

scales network depth, width and input resolution jointly instead of separately, which gives it a better accuracy-to-compute trade-off than earlier architectures — a property that matters when computational resources are limited, as is often the case in medical imaging applications. Lightweight architectures such as MobileNets [5] pursue a similar efficiency goal through depthwise separable convolutions; EfficientNet's compound scaling, however, tends to reach higher accuracy at a comparable computational budget, which is part of why we adopt it here rather than a MobileNet-style backbone.

2.2. Metric Learning and ArcFace Loss

Metric learning aims to train an encoder that maps inputs into a latent space where semantically similar examples end up close together. ArcFace, introduced by Deng et al. [6] for face recognition, adds an angular margin penalty to the softmax loss, which forces a larger angular separation between classes than standard softmax cross-entropy would. Contrastive approaches such as SimCLR [7] and supervised contrastive learning [8] pursue a related idea by pulling embeddings of similar examples together and pushing dissimilar ones apart, but without the explicit angular margin that ArcFace enforces between class centers.

2.3. Unsupervised Clustering of Deep Embeddings

K-Means [9], a classical partitioning method covered in most standard machine learning texts [10], is still one of the most widely used clustering algorithms, mainly because it is simple and cheap to run. Its main drawback is that the number of clusters k has to be fixed in advance, and it is sensitive to initialization and to the assumption of roughly spherical, convex clusters. HDBSCAN [11] takes a different route: it builds a condensed hierarchical tree of the data and keeps the clusters that are the most persistent according to a stability criterion, without ever needing k as an input. This is particularly convenient in unsupervised medical image analysis, where the true number of diagnostic categories is not always known beforehand. For visualizing the resulting high-dimensional embeddings, UMAP [12] has become a common choice in biological and clinical image analysis. More broadly, our pipeline fits within a growing body of self-supervised and label-efficient learning methods [13] that try to reduce the dependency of medical imaging models on large, manually annotated datasets.

3. Proposed Methodology

3.1. Dataset and Pre-processing

We use a six-class lung cancer histopathology dataset of 3,174 images from a public Kaggle repository, covering BenignCases, MalignantCases, adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal tissue. Two pre-processing configurations are evaluated: a balanced version, where all classes are equalized through oversampling with data augmentation, and the original imbalanced version, which keeps the natural class frequencies. Pre-processing is implemented in preprocess_img.py and includes resizing to 224×224 pixels, normalization, and augmentation (random flipping, rotation, brightness/contrast jittering, and Gaussian blur).

3.2. EfficientNetB0 with ArcFace

EfficientNetB0, initialized with ImageNet pre-trained weights, serves as the backbone feature extractor. We replace its classification head with a dense embedding layer that outputs 512-dimensional representations, followed by an ArcFace classification head applying an additive angular margin penalty to the normalized embedding-to-class-center cosine distances. The ArcFace loss is defined as:

$$L = -log\left[\frac{e^{s\cdot cos(\theta_{yi}+m)}}{e^{s\cdot cos(\theta_{yi}+m)} + \sum_{j\neq yi} e^{s\cdot cos(\theta_j)}}\right]$$

where s is the scale factor, m the angular margin penalty, and θ_{yi} the angle between the normalized embedding and the weight vector of the true class i. This formulation angularly separates classes by at least a fixed margin, which pushes the resulting representation space toward strong inter-class discriminability.

We opted for ArcFace over these contrastive alternatives because its margin-based formulation gives a more direct control over inter-class separability, which is precisely the property our downstream clustering step relies on. In practice, this produces embeddings that stay discriminative even when they are later handed to an unsupervised algorithm with no label information — which is exactly what makes ArcFace attractive for a hybrid supervised-unsupervised pipeline like ours.

3.3. Three-Phase Progressive Training

Training is split into three progressive phases to help convergence and limit overfitting. In Phase 1, the EfficientNetB0 convolutional weights stay frozen and only the ArcFace head is trained, which brings validation accuracy from around 30% up to about 88% fairly quickly. Phase 2 gradually unfreezes the upper convolutional blocks, allowing the network to adapt to the domain and pushing validation accuracy to roughly 97%. Phase 3 fine-tunes every weight jointly with a reduced learning rate, reaching a best validation accuracy of 98.19%. The whole process runs for 100 epochs with StratifiedKFold splits, early stopping on validation loss, and checkpointing at the best validation accuracy.

3.4. Unsupervised Clustering Strategy

Once training is finished, we extract 512-dimensional embeddings for every sample and cluster them. K-Means is run with k = 6, matching the known number of classes; HDBSCAN is run without specifying k at all, relying purely on the density structure of the embedding space to find the groups. Cluster-to-class alignment for accuracy computation uses the Hungarian algorithm. Clustering quality is assessed under 10-fold stratified cross-validation using five metrics: clustering accuracy (ACC), Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), F1-score, and, for HDBSCAN, the noise rate.

4. Experimental Setup

4.1. Implementation Details

All experiments are implemented in Python 3.10 with TensorFlow/Keras 2.x and scikit-learn. EfficientNetB0 weights come from the Keras Applications module. ArcFace is implemented as a custom Keras layer with margin m = 0.5 and scale s = 64. HDBSCAN is run with a minimum cluster size of 15 and the Euclidean metric on the full 512-dimensional embedding space, and K-Means uses k-means++ initialization with 10 random restarts. All runs were carried out on a workstation with a standard GPU.

4.2. Evaluation Protocol

Performance is evaluated with 10-fold stratified cross-validation, so that each fold keeps the original class distribution. We evaluate two configurations: balanced (oversampled with augmentation) and imbalanced (original class frequencies). For every fold, the model is retrained from scratch to avoid any information leakage between folds. Statistical significance between K-Means and HDBSCAN is checked with a paired two-tailed t-test, using $p < 0.05$ as the significance threshold.

5. Experimental Results

5.1. Supervised Classification — Balanced Dataset

The three-phase training reaches a best validation accuracy of 98.19% and a test accuracy of 98.10% (773 correct out of 787 images). Training and validation loss both decrease monotonically across all three phases, from around 22.0 down to nearly 0.0, and the two curves stay closely aligned throughout — a good sign that the model is not overfitting. We attribute this mostly to the progressive training schedule and to the implicit regularization that comes from optimizing the ArcFace angular margin.

Table 1 breaks down performance by class on the test set. MalignantCases is classified perfectly (100%), while large cell carcinoma has the lowest precision (95.2%), with most of its errors going to adenocarcinoma — which is not surprising biologically, since the two subtypes share overlapping morphological features at the cellular level.

Table 1. Per-class classification performance on the test set (balanced dataset, accuracy = 98,00%).

Class	Correct	Total	Precision	Errors
BenignCases	97	99	97.9%	2 → normal
MalignantCases	169	169	100%	None
adenocarcinoma	146	152	96.0%	3 → large cell, 3 → squamous
large cell carcinoma	80	84	95.2%	4 → adenocarcinoma
normal	164	168	97.6%	Minor confusions
squamous cell carcinoma	115	117	98.3%	2 → large cell

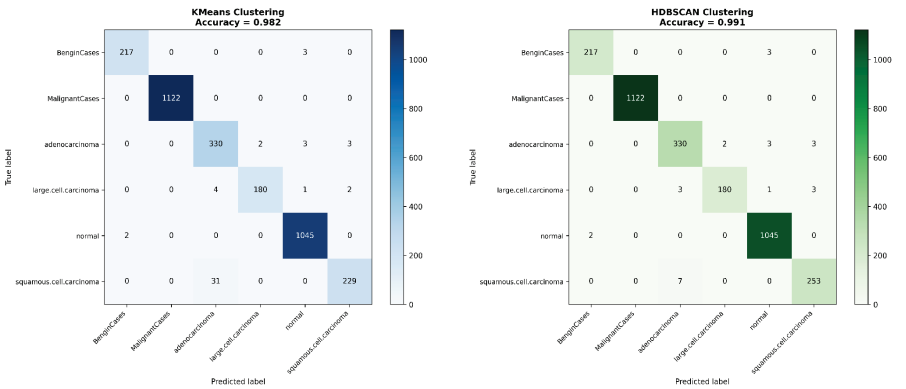


Figure 1. Aggregate confusion matrices for k-Means (left) and HDBSCAN (right) over 10-fold cross-validation — balanced dataset. Both algoriths reach 99.8% accuracy with near-identical error patterns.

5.2. Embedding Space — UMAP Visualization

The UMAP 2D projection of the 512-dimensional embedding space shows six compact, well-separated clusters, one per class, with no visible overlap between them. This is a direct, visual confirmation that ArcFace training does what it is supposed to do — enforce an angular margin in the embedding space — and it also explains why the downstream clustering works so well. Interestingly, even adenocarcinoma and large cell carcinoma, which look visually similar under the microscope, end up in clearly distinct regions of the UMAP space, suggesting the 512-dimensional representation captures more discriminative detail than what a 2D projection alone can show.

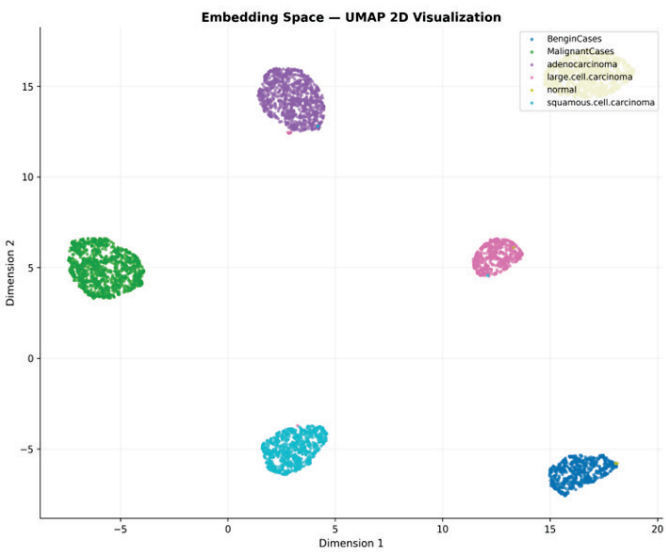


Figure 2. UMAP 2D projection of the 512-dimensional ArcFace embedding space. All six lung cancer classes form compact, non-overlapping clusters

5.3. Unsupervised Clustering — Balanced Dataset

Under 10-fold stratified cross-validation on the balanced dataset, K-Means and HDBSCAN both reach a clustering accuracy of about 99.8%, and their scores are close on every metric we tracked. Table 2 summarizes the comparison.

Table 2. Clustering performance comparison — balanced dataset (10-fold cross-validation)

Metric	K-Means	HDBSCAN	Significance
Accuracy	~99.8%	~99.8%	ns ($p > 0.05$)
F1-score	~99.8%	~99.8%	ns
ARI	~0.994–0.996	~0.994–0.996	ns
NMI	~0.993	~0.993	ns
Noise (%)	N/A	0.0%	—
Clusters detected	6 (fixed)	6 (automatic)	Exact match

What stands out most is that HDBSCAN, which is never told how many classes exist, still recovers the six-cluster structure with 99.8% accuracy across all ten folds, a stable ARI around 0.996, and zero noise points. That result is hard to get by chance: it tells us the ArcFace embeddings form dense, well-separated structures in the full 512-dimensional space, structures that a density-based method can find reliably on its own.

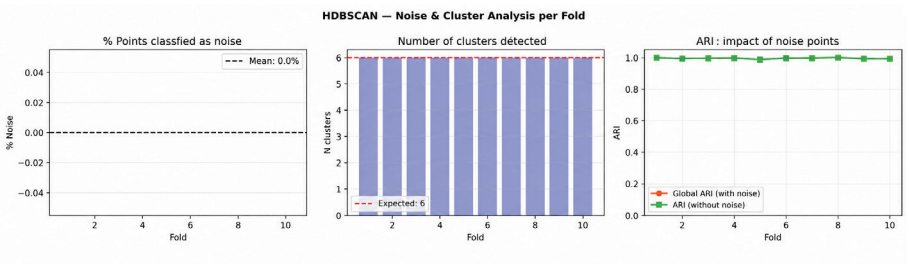


Figure 3. HDBSCAN noise and cluster analysis per fold (balanced dataset). Left : zero noise across all 10 folds. Center: exactly 6 clusters detected automatically at each fold. Right: ARI stable at ~0.996 throughout cross-validation.

5.4. Effect of class Imbalance

To check how robust the pipeline is under more realistic conditions, we repeat the experiments on the original, imbalanced dataset. Table 3 reports the main differences

Table 3. Impact of class imbalance on clustering performance.

Aspect	Imbalanced	Balanced	Difference
Accuracy – K-Means	98.4%	99.8%	−1.4%
Accuracy – HDBSCAN	99.1%	99.8%	−0.7%
NMI	~0.970–0.975	~0.993	+0.02 (balanced)
Cross-fold variance	High (outliers)	Low (stable)	—
squamous/adeno errors	31 images (K-Means)	1–2 images	~15x reduction
UMAP structure	Spread clusters	Compact clusters	—

The clearest sign of degradation under imbalance shows up for squamous cell carcinoma with K-Means: 31 images end up misclassified as adenocarcinoma, a 12% class-level error rate. HDBSCAN cuts that down to 7 misclassifications, which speaks to its robustness against uneven inter-class density. HDBSCAN still reports zero noise points even in this imbalanced setting, so the embedding clusters keep their structural integrity regardless of class frequency. The paired t-test between K-Means and HDBSCAN metrics is not significant on the balanced dataset, but on the imbalanced one HDBSCAN shows a consistent, practical edge — especially in cutting down critical errors on minority classes.

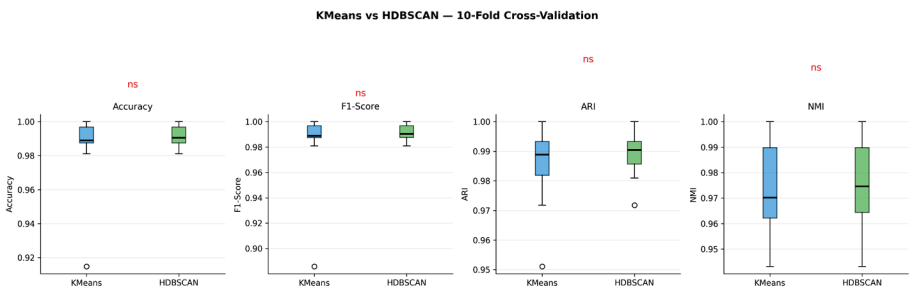


Figure 4. Boxplot comparison of K-Means vs. HDBSCAN over 10-fold cross-validation on the imbalanced dataset. K-Means shows outlier folds (~92%) that are absent for HDBSCAN; all pairwise differences remain non-significant (ns, $p > 0.05$).

6. Discussion

6.1. Interpretation of Results

Taken together, these results show that combining EfficientNetB0 with an ArcFace loss is a genuinely effective embedding strategy for lung cancer histopathology. The 98.10% supervised accuracy, obtained on a strict train/test split with no data leakage, is competitive with the best approaches reported for this type of problem. What is more striking, though, is the near-perfect unsupervised clustering accuracy of 99.8%: it shows that the ArcFace embeddings carry enough structure to support label-free diagnosis — a finding that could matter for clinical deployment in settings where annotated data is hard to come by.

The progressive, three-phase training schedule seems essential to reaching this outcome. Freezing the backbone in Phase 1 lets the ArcFace head settle into a stable angular geometry before the lower-level features are disturbed by fine-tuning. Training and validation curves stay close throughout all three phases, with no sign of a growing gap between them, which points to solid generalization despite the high dimensionality of the embedding space.

The HDBSCAN noise analysis is, in our view, one of the more convincing indicators of embedding quality here. Finding exactly six clusters with zero noise across all ten folds is not something that happens with a merely approximate structure — it means the embedding space forms dense, well-separated manifolds that line up precisely with the six biological categories. That property could be clinically useful too, since noise rejection could flag ambiguous cases automatically for a specialist to review.

6.2. Comparison with Related Work

Two recent studies give a useful point of comparison, since both use the same Kaggle CT-scan dataset and the same four overlapping classes (adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal tissue). Shatnawi et al. [14] proposed a modified DenseNet201-based CDL model with feature selection and reported 93.24% accuracy on this dataset; the paper does not report a single, unambiguous precision/recall/F1 breakdown for that specific model, since several architectures were compared side by side, so we limit the comparison to overall accuracy for this study. Huang et al. [15] fine-tuned a ResNet50V2 backbone with Squeeze-and-Excitation attention blocks and reported a more complete picture: 90.16% test accuracy, an overall AUC of 0.9815, per-class precision ranging from 0.81 (large cell carcinoma) to 1.00 (normal tissue), recall from 0.85 (adenocarcinoma) to 0.98 (large cell carcinoma), and F1-scores between 0.88 and 0.96.

Our pipeline reaches 98.10% supervised test accuracy on a superset of six classes, which is 4.86 points above [14] and 7.94 points above [15], despite covering two additional diagnostic categories (BenignCases and MalignantCases) rather than fewer. The class-level picture is informative too: large cell carcinoma is the weakest class in both studies, but where Huang et al. [15] report only 0.81 precision for it, our model reaches 95.2% precision on the same class (Table 1) — a 14-point gap on exactly the class that both pipelines struggle with the most, which suggests the ArcFace-based embedding separates morphologically similar subtypes more cleanly than a standard softmax-trained ResNet50V2. We should note two caveats for fairness: [14] and [15] both train on roughly 900–1,000 images, about a third of our 3,174-image dataset, and their four-class setup is a simpler problem than our six-class one, so part of the accuracy gap may also reflect the larger, more diverse training set available to us rather than the architecture alone.

Beyond the raw accuracy gap, neither [14] nor [15] explores an unsupervised classification pathway; both rely entirely on supervised cross-entropy training. Our results additionally show that the same ArcFace embeddings support near-perfect unsupervised clustering (about 99.8% accuracy), which is not something a standard softmax-trained CNN like theirs is designed to provide. As far as we know, ArcFace has not been widely applied to lung histopathology before, which suggests that metric-learning losses developed for biometrics can transfer reasonably well to clinical imaging.

6.3. Limitations

A few limitations are worth flagging. First, the dataset, though public and widely used as a benchmark, may not capture all the variability seen in clinical practice across institutions and staining protocols. Second, we did not measure inference latency, which matters for any real-time diagnostic tool. Third, our evaluation stays at the image level and does not address spatial localization or the interpretability of the learned features. Finally, the ArcFace margin hyperparameters were kept at the standard values used in face recognition rather than tuned specifically for histopathological embeddings.

7. Conclusion

We presented a deep learning pipeline for six-class lung cancer image classification that combines EfficientNetB0 with ArcFace metric learning and density-based unsupervised clustering. The pipeline reaches 98.10% supervised accuracy and roughly 99.8% unsupervised clustering accuracy under 10-fold stratified cross-validation on 3,174 histopathology images. HDBSCAN recovers all six diagnostic categories on its own, with zero noise across every validation fold, which reflects the strong separability of the learned 512-dimensional embedding space also visible in the UMAP projection. Looking at class imbalance, oversampling with augmentation improves every metric we tracked and substantially cuts down critical minority-class errors, with roughly a 15-fold drop in squamous/adenocarcinoma confusion.

These results suggest that metric-learning-enhanced deep embeddings are a solid basis for both supervised and unsupervised lung cancer diagnosis. The pipeline is a practical, reproducible step toward label-efficient clinical decision support in pulmonary pathology.

Acknowledgements

The authors thank the Faculty of Sciences and Technology of the University of Toamasina for its support throughout this research.

Conflicts of Interest

The authors declare no conflict of interest regarding the publication of this work.

References

[1] Emmerick, I.C.M., Singh, A., Powers, M., Lou, F., Lin, P., Maxfield, M. and Uy, K. (2021) Factors Associated with Diagnosis of Stages I and II Lung Cancer: A Multivariate Analysis. Revista de Saúde Pública, 55, 112. https://doi.org/10.11606/s1518-8787.2021055003345

[2] Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning. MIT Press, Cambridge. https://www.deeplearningbook.org/

[3] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. Proceedings of the IEEE/CVF CVPR, 770-778. https://arxiv.org/abs/1512.03385

[4] Tan, M. and Le, Q. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the International Conference on Machine Learning (ICML), 6105-6114. https://arxiv.org/abs/1905.11946

[5] Howard, A.G., Zhu, M., Chen, B., et al. (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861. https://arxiv.org/abs/1704.04861

[6] Deng, J., Guo, J., Xue, N. and Zafeiriou, S. (2019) ArcFace: Additive Angular Margin Loss for Deep Face Recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4690-4699. https://arxiv.org/abs/1801.07698

[7] Chen, T., Kornblith, S., Norouzi, M. and Hinton, G. (2020) A Simple Framework for Contrastive Learning of Visual Representations. Proceedings of the International Conference on Machine Learning (ICML). https://arxiv.org/abs/2002.05709

[8] Khosla, P., Teterwak, P., Wang, C., et al. (2020) Supervised Contrastive Learning. Advances in Neural Information Processing Systems (NeurIPS). https://arxiv.org/abs/2004.11362

[9] Lloys, S. (1982) Least Squares Quantization in PCM. IEEE Transactions on Information Theory, 28, 129-137. https://doi.org/10.1109/TIT.1982.1056489

[10] Bishop, C.M. (2006) Pattern Recognition and Machine Learning Springer, New York. https://link.springer.com/book/9780387310732

[11] Campello, R.J.G.B., Moulavi, D. and Sander, J. (2013) Density-Based Clustering Based on Hierarchical Density Estimates. Proceedings of the Pacific-Asia conference on Knowledge Discovery and Data Mining (PAKDD), 160-172. https://doi.org/10.1007/978-3-642-37456-2_14

[12] McInnes, L., Healy, J. and Melville, J. (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426. https://arxiv.org/abs/1802.03426

[13] Jing, L. and Tian, Y. (2020) Self-Supervised Visual Feature Learning with Deep Neural Networks: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://arxiv.org/abs/1902.06162

[14] Shatnawi, M.Q., Abuein, Q. and Al-Quraan, R. (2024) Deep Learning-Based Approach to Diagnose Lung Cancer Using CT-Scan Images. Intelligence-Based Medicine, 11, Article 100188. https://doi.org/10.1016/j.ibmed.2024.100188

[15] Huang, K.A., Venkitasubramony, V. and Prakash, N.S. (2025) Leveraging Transfer Learning and Attention Mechanisms for a Computed Tomography Lung Cancer Classification Model. Cureus, 17, e87071. https://doi.org/10.7759/cureus.87071