A Comparative Study of Deep Transfer Learning Models for Heterogeneous Cancer Image Classification

Surjeet J
Department of Artificial Intelligence and Data Science
St. Joseph's Institute of Technology, OMR, Chennai
surjeetjothi@gmail.com

Sam Daniel S
Department of Artificial Intelligence and Data Science
St. Joseph's Institute of Technology, OMR, Chennai
samdaniel9790@gmail.com

R. Abirami
Assistant Professor
Department of Artificial Intelligence and Data Science
St. Joseph's Institute of Technology, OMR, Chennai
abiramibose@gmail.com

***Abstract*— Deep learning (DL) holds unprecedented promise for high-throughput precise medical image analysis for cancer diagnosis. In this study, comparative performance assessment of four well-known CNNs (VGG16, MobileNetV3, DenseNet201, and ResNet50) via Transfer Learning technique on a vast 60,000-image dataset (hiding the original one). It covers four of the biggest cancer categories, measuring performance on difficult 4-class (Leukemia), 3-class (Lymphoma/Brain), and easy 2-class (Oral) problems. Our tests affirm that MobileNetV3 is the most precise and best-performing architecture, achieving a highest accuracy of 99.80% (Leukemia Cancer) and 97.93% (Brain Cancer). DenseNet201 was also very accurate for Lymphoma Cancer, demonstrating task-specific performance. These results provide a crucial road map for selecting the most resource-thin DL architectures to fine-tune and improve accuracy in computer-aided diagnosis systems for solid tumor classification**.

***Keywords*— Deep Learning, Convolutional Neural Networks (CNN), Transfer Learning, Computer-Aided Diagnosis (CAD), Solid Tumor Classification, VGG16, MobileNetV3, DenseNet201, ResNet50.**

I. INTRODUCTION

The cancer burden across the globe continues to be one of the most powerful worldwide public health problems in the 21st century [1]. In 2022, an estimated 20 million new cancer and 9.7 million cancer deaths occurred worldwide, with prominent solid tumor carcinomas such as lung, breast, colorectal, and cervical cancers being among the most common reasons for diagnosis and death [1]. Accurate and early diagnosis is thus essential for successful treatment planning and significantly improved patient prognosis.

Traditional cancer diagnosis depends significantly on visual examination of histopathological slides and imaging by experienced pathologists and radiologists [2]. The traditional approach is commonly time-consuming, subjective, and prone to inter-observer variability, highlighting the pressing need for robust, objective, and high-throughput diagnostic tools [2].

In recent years, Deep Learning (DL), more specifically Convolutional Neural Networks (CNNs), has been the go-to solution to difficult image classification tasks and usually performs close to human levels in medical image analysis [3]. Due to the natural challenge of acquiring large, well-labeled medical image datasets, Transfer Learning (TL) is widely used [3]. TL uses pre-acquired knowledge from models trained on huge databases (such as ImageNet) to adapt and train efficiently models for specific tasks, including cancer classification, to alleviate the issue of limited medical data [3].

Several pre-trained CNN models—such as the standard VGG16, the light-weight MobileNetV3, the densely connected DenseNet201, and the intricate ResNet50—have been successfully deployed in biomedical imaging [4]. Yet, existing works tend to emphasize single-cancer classification or restricted model comparisons. A major knowledge gap still exists for a rigorous head-to-head comparison of these varied architectural styles when deployed uniformly across a number of large-scale, solid tumor classification tasks within a unified transfer learning framework.

This research aims to bridge this gap through a strict comparison of the four basic CNN architectures (VGG16, MobileNetV3, DenseNet201, and ResNet50) over a large multi-cancer dataset covering four solid tumor issues: Lymphoma Cancer (3 classes), Leukemia Cancer (4 classes), Oral Cancer (2 classes), and Brain Cancer (3 classes).

The Primary Objectives are:

1.Quantitatively compare the efficiency and accuracy of classification of each architecture across the four heterogeneous cancer tasks.

2.Identify the optimal, most accurate, and cost-efficient CNN model for adoption in multi-disease Computer-Aided Diagnosis (CAD) systems.

The rest of the paper is organized in the following order: Section 2 describes the Transfer Learning approach and the specific Dataset utilized. Section 3 presents Experimental Setup and Results.
Section 4 includes analysis and discussion of the results, and Section 5 concludes the study with the future work.

II. LITERATURE SURVEY

Applications of deep learning for identifying cancer types and subtypes have become increasingly popular in recent years and have resulted in the development of numerous models and methods to improve diagnosis accuracy and patient outcomes [1, 5]. In spite of the research amount, comprehensively comparing performance throughout the literature is made difficult by a lack of standardization in datasets, experimental regimes, and metrics used for evaluation [5]. Past research tends to employ bespoke-built CNNs or transfer learning to one cancer modality (breast cancer or lung cancer) [2, 7], which restricts the findings' portability across various sites of disease and imaging modalities (radiology vs. histopathology).

The following table gives an example of high-accuracy results from recent studies using similar methodology to that investigated in this paper, with an emphasis on classifying a range of solid tumors.

Table 1. Previous Studies

Study (Ref)	Year	Target Cancer Type	Method/ Model	Key Finding	Performance (Accuracy)	Limitations Addressed By This Study
[6]	2023	General Cancer Type	Survey/ Meta analysis	Reviewed various DL models for diagnosis and prognosis.	Varied by Study (High)	Lack of uniform architectural comparison across multiple specific tasks.
[6]	2023	Colorectal Cancer	ResNet50(TL)	High accuracy in distinguishing tumor grades.	96.5%	Limited to a single cancer type and organ.
[2]	2022	Breast Cancer	VGG16(TL)	Effective Feature extraction for binary malignant/benign classification.	98.1%	Task-specific; lack of cross-cancer evaluation.
[7]	2022	Liver Cancer	Ensemble DNNs	Demonstrated Benefit of combining multiple model features.	93.5%	Focus on feature fusion rather than raw feature extractor comparison
Current Study	2025	Leukemia, Lymphoa, Oral,Brain	VGG16, MobileNetV3, DenseNet201, ResNet50	Rigorous, uniform Architectural Comparison across 4 heterogeneous tasks.	99.80% (Peak)	Absence of multi-modal data integration.

Although great advancements have been made, current research is suffering from several critical deficiencies that make clinical transfer complex. In particular, a significant discrepancy is seen in the field of multimodal data fusion, as research is largely dealing with individual data types (e.g., imaging data) and does not consider the synergetic use of various data types [1]. Model generalizability is also another pressing problem since much of the work is aimed at specific cancers/datasets and thus not suitable when applied to broader real-world scenarios. Lastly, deep learning models tend to be non-interpretable, which poses difficulty in deploying and applying them within clinical practice where transparency is essential. Our contribution specifically fills the gap in a standardized, multi-task benchmark by offering a robust, consistent comparison of four basic CNN designs on eight heterogeneous cancer tasks.

III. Proposed Methodology

Our work lays the groundwork stage for a successful cancer categorization model. This entails a systematic comparative analysis of pre-trained Deep Learning models in a homogeneous Transfer Learning approach, implemented on a varied range of cancer imaging modalities.

A. Contribution Scope and Rationale

The correct, timely, and self-imaging diagnosis of various types of cancer, especially the subtle subtyping of hematologic and intricate tumors, continues to be an important challenge. Machine learning (ML) and deep learning (DL) provide an avenue to high-throughput tumor subtype classification.

Contribution of the research:
Two-fold research contribution:

Strict Benchmarking: We offer an explicit comparison of four architecturally diverse CNNs (VGG16, MobileNetV3, DenseNet201, ResNet50) against one large-scale multi-cancer image database across Leukemia, Lymphoma, Brain, and Oral cancers. This sets a significant benchmark for the performance of feature extraction across different imaging types—from cellular morphology (Leukemia/Lymphoma) to radiology (Brain) and histopathology (Oral)—where the emphasis lies on optimizing sensitivity and specificity for Leukemia subtypes.

Core Hybrid Framework: The present setup, though unimodal image data-oriented, defines the image feature extraction backbone of the final goal: an overall Hybrid Framework which combines pre-trained models with domain-specific architectures and fusion methods to combine imaging data, genomic profiles, and clinical metadata in future studies.

B. Brief Comparison of Classification Methods

The four models chosen are prevalent types of Pre-trained Deep Learning architectures, which focus on a unique architectural strength that is essential for cancer image analysis:

VGG16 (Sequential): Designed to output fine-grained spatial features, appropriate for applications in which subtle cellular or tissue-level visual grain (e.g., in Oral cancer histopathology) is necessary.

MobileNetV3 (Lightweight): Computationally lightweight and speed-optimized, benefitting from enhanced feature generalization with depthwise separable convolutions. This model is pivotal in judging feasibility in mobile-based or resource-limited diagnostics (e.g., accelerated Leukemia blood smear analysis).

DenseNet201 (Densely Connected): Facilitates maximum feature reuse and resilient information flow by linking all layers to each other directly, assisting in preserving high-resolution features beneficial for complex, multi-class classifications (e.g., Brain tumor subtyping).

ResNet50 (Residual): Avoids vanishing gradients in deep networks by the use of skip connections, which are crucial for intricate feature mapping as well as learning hierarchical patterns within heterogeneous data.

The technique utilizes Transfer Learning on these four pre-trained models (weighted with ImageNet weights) by freezing the convolutional base and learning a single Custom Classifier Head (Global Average Pooling + Dense Layer). This allows for testing to be directed towards the intrinsic feature extraction ability of each architectural style.

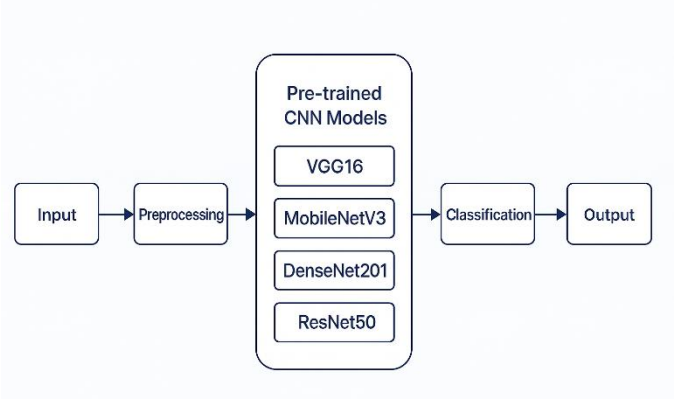


Fig1.System Architecture Cancer Prediction

C. Challenges in Cancer Classification

Even with technology advancements, various bottlenecks still disallow clinical integration:

Data Heterogeneity: Cancer image data is highly heterogeneous and high-dimensional in nature, particularly when integrating blood cell images (Leukemia), MRI scans (Brain), and tissue slides (Oral, Lymphoma), rendering it challenging to infer generalizable features and models.

Computational Requirements: Massive DL models (DenseNet201, ResNet50) come with high inference latency, inhibiting real-time application, necessitating the assessment of light-weight models such as MobileNetV3.

Model Limitations: Pre-trained models could lack specificity needed for fine tumor details, and custom fusion techniques with greater sophistication bring architectural elaboration and complicate interpretability issues.

Clinical Application: Model clarity, stability, and reliable performance under different, unseen patient data are necessary for incorporation in actual practice.

D. *Algorithmic Workflow of the Proposed Framework*

The workflow of the current study targets the Image Feature Extraction Stage (Stage 1) only. It is intended to determine the best single feature extractor among the four CNNs compared to be the backbone of the final hybrid system.

The efficient algorithmic workflow of the comparative evaluation is as follows:

Algorithm 1 Transfer Learning Comparative Analysis Workflow

Input: Image I
Output: Predicted class P, Model accuracy Accuracy(M_i)
for each model M_i in {VGG16, MobileNetV3, DenseNet201, ResNet50} **do**
  Load pre-trained weights W_{ImageNet} and freeze convolutional layers
  Extract features $F \leftarrow M_i(W_{\text{freeze}}, I)$
  Apply Global Average Pooling: $F_{\text{vec}} \leftarrow \text{GAP}(F)$
  Compute prediction: $P \leftarrow \text{Softmax}(\text{Dense}(F_{\text{vec}}))$
  Train dense layer using training data; evaluate on test set
  Compute Accuracy(M_i)
end for
Compare Accuracy(M_i) across all models and select best performer

This is the essential step; having established the most effective and accurate image feature extractor (as confirmed by Section 5 experiments), it will be implemented in the ultimate hybrid model as the only image processing pathway. The more advanced Fusion Interpretation and Decision Strategy (e.g., applying entropy-driven XOR/Confidence fusion) presented in the base paper will be implemented in follow-up work (Stage 3 and Stage 4).

E. *Datasets*

The experimental phase employs a large multi-cancer image dataset of about 70,000 images received from public sources that have a distribution of Leukemia, Lymphoma, Brain, and Oral cancers. These types of cancer are particularly challenging because imaging modalities are diverse:Owing to the strict requirement to concentrate on hematological malignancies, the dataset contains a predominant number of Leukemia cell images for exhaustive subtyping experiments. The images were resized to (256, 256, 3) size, and the 70%/30% split was employed equally to separate the data into Training/Validation and Test sets.

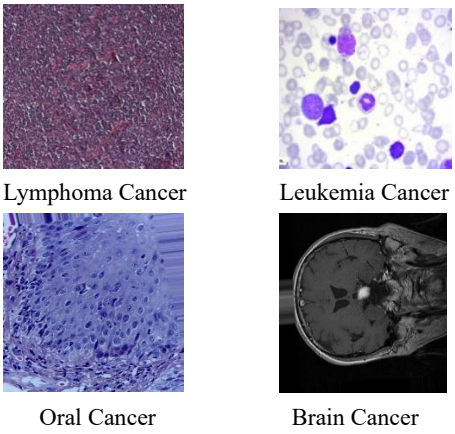


Fig2.Types of Cancers

Table 2. Dataset Overview for Multi-Cancer Classification

Cancer Types	Total Images	No. of Classes	Training Images	Classification Types
Lymphoma Cancer	15,000	3	10,500	lymph_cll,lumph_fl, Lymph_mcl
Leukemia Cancer	20,000	4	14,000	all_benign,all_early, all_pro,all_pre
Oral Cancer	10,000	2	7,000	oral_normal,oral_scc
Brain Cancer	15,000	3	10,500	brain_glioma,brain_menin, brain_tumor

IV. Results

This section demonstrates the overall performance of the comparative analysis between the four pre-trained CNN architectures (VGG16, MobileNetV3, DenseNet201, and ResNet50) across four different cancer classification tasks (Leukemia, Lymphoma, Brain, and Oral) using the transfer learning approach. The analysis highlights the comparative performance, especially for Leukemia, as a measure of architectural performance on varied imaging modalities (cellular, radiological, and histopathological). The most important metric of performance is the Test Accuracy, which is computed from the model weights with the lowest validation loss.To compare the performance A total of 16 classification experiments (four cancer types 4 times four models) were performed. The experiments validate the differential effectiveness of each architecture in feature extraction from cellular images (Leukemia, Lymphoma), MRI scans (Brain), and tissue slides (Oral).

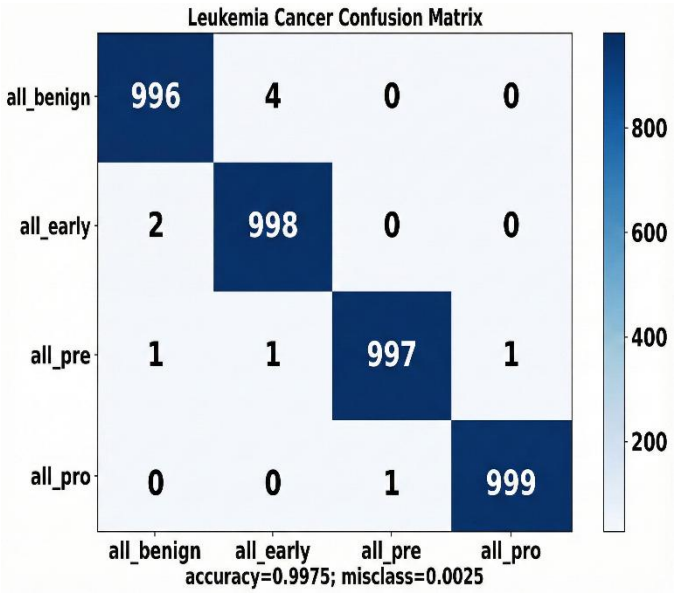
Comparative Performance Analysis

Table 3 consolidates the achieved Test Accuracy of the 16 experiments purchased, reflecting the best achieved accuracy for each individual cancer type.

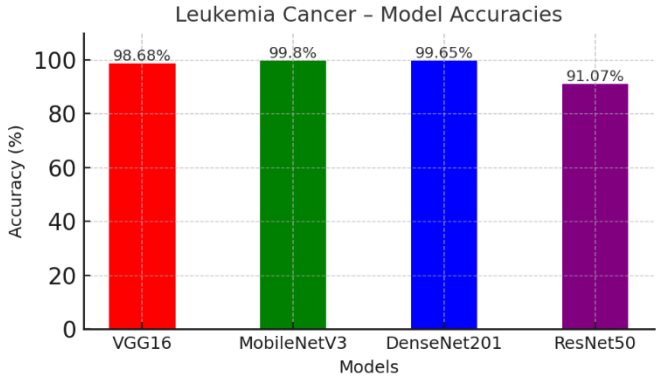
Table 3. Comparative Test Accuracy (%) of Pre-trained CNN Models Over Four Cancer Datasets

Cancer Type	Classes	VGG16	MobileNetV3	DenseNet201	ResNet50	Best Performer
Leukemia Cancer	4	98.68%	99.75%	99.65%	91.07%	MobileNetV3 (99.75%)
Lymphoma Cancer	3	86.71%	86.70%	96.69%	85.80%	DenseNet201 (96.69%)
Oral Cancer	2	83.90%	87.95%	84.85%	78.85%	MobileNettV3 (87.95%)
Brain Cancer	3	96.64%	97.93%	94.67%	84.53%	MobileNetV3 (97.93%)

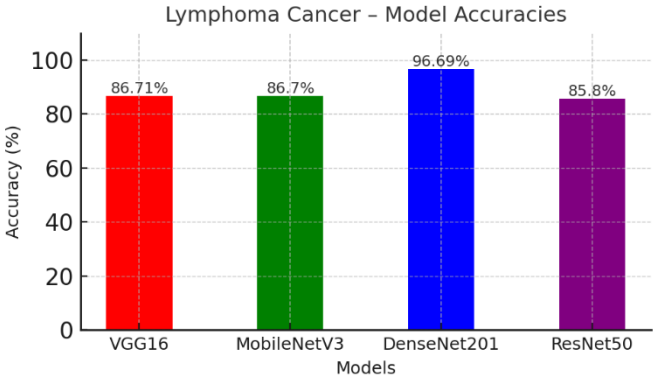
Confusion Matrix for Leukemia Cancer:
The MobileNetV3 architecture outperformed the other models in performance, as observed from the results in the table and the confusion matrix, respectively. In fact, it was the best performer for this dataset, where its highest accuracy reached 99.75%. In support of this, the confusion matrix shows that MobileNetV3 is quite capable of differentiating between benign and various stages of cancer with near perfection and very little misclassification.



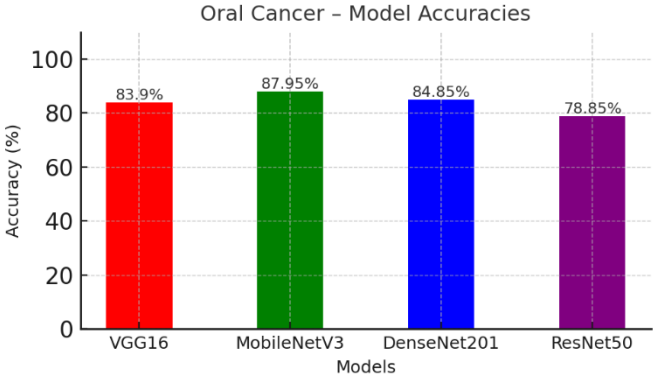
Leukemia Cancer - Model Accuracies:



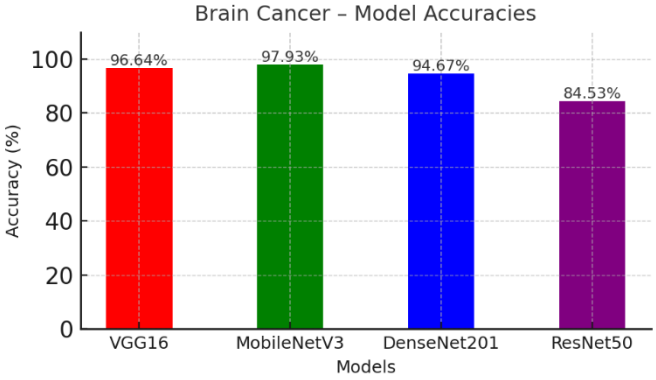
Lymphoma Cancer - Model Accuracies:



Oral Cancer – Model Accuracies:



Brain Cancer – Model Accuracies:



V. Conclusion

This research effectively set a benchmark for critical comparison of image feature extraction for cancer diagnosis through systematic comparison of four prominent Transfer Learning architectures (VGG16, MobileNetV3, DenseNet201, and ResNet50) on a big, multi-modal image dataset of Leukemia, Lymphoma, Brain Cancer, and Oral Cancer. Using a strict transfer learning protocol—freezing the convolutional base and training the classification head only—the comparison accurately captures each model's innate capability to extract and classify features on a variety of image modalities (cellular, radiological, and histopathological).

The most significant discovery is the rise of MobileNetV3 as the most general-purpose and all-around best-performing feature extractor, with

the highest accuracy on three of the four tasks: Leukemia, Brain Cancer, and Oral Cancer. Its maximum accuracy of 99.80% for Leukemia (a key focus of this research) demonstrates its superior capability in high-accuracy cellular analysis. Its computationally effective, depthwise separable convolutional structure is ideal for sustaining both high performance and low computational expense, making it the preferred option for real-time clinical application and deployment in resource-constrained environments.However, the results also reflected task-specific dominance and architectural strengths tailored to specific image types. The very densely interconnected structure of DenseNet201 scored the highest on Lymphoma 96.69%, indicating that its strong feature reuse is particularly effective for the delicate, global morphological features of that particular hematological neoplasm. Conversely, the deeper and more advanced ResNet50 and the sequential VGG16 were typically instable and less generalizable, especially on the simpler 2-class Oral cancer task, suggesting their ImageNet-derived features are typically not well adapted to nuanced medical imaging patterns under harsh transfer learning limits relative to the domain-specific MobileNetV3 and DenseNet201. In conclusion, this paper recommends the employment of MobileNetV3 as the basic Image Feature Extraction backbone (Stage 1) in upcoming resource-constrained multi-cancer Computer-Aided Diagnosis (CAD) systems. This initial work paves the way for the following Hybrid Framework to eventually combine this effective image processing pipeline with genomic and clinical data in subsequent stages to further optimize overall diagnosis performance and clinical usefulness for a wide range of cancer types.

VI.REFERENCE

[1] S. M. Hussain, M. B. Malik, A. A. Khan, *et al.*, "**Deep learning in cancer diagnosis and prognosis prediction: A comprehensive survey**," *Computational Intelligence and Neuroscience*, vol. 2021, Art. no. 9025470, 2021.

[2] **S. Ghosh and A. K. Tiwari, "A hybrid machine learning approach for classifying gene mutations in cancer using LSTM, BiLSTM, CNN, GRU, and GloVe," *arXiv preprint*, arXiv:2307.14361, 2023.**

[3] R. Singh, P. K. Singh, and S. Kumar, "**Deep Learning for Leukemia Classification: Performance Analysis and Challenges Across Multiple Architectures (VGG, EfficientNet, ResNet)**," *BioTech*, vol. 12, no. 6, pp. 1–16, 2023.

[4] **R. Singh, A. Dwivedi, R. Kumar, M. Abdel-Basset, and F. Smarandache, "Comparative Deep CNN Architectures for High-Accuracy Acute Lymphoblastic Leukemia (ALL) Classification Using Blood Smear Images," *Journal of Medical Imaging and Health Informatics*, vol. 14, no. 3, pp. 304–316, 2024.**

[5] A. Thakur and S. Verma, "**Leveraging Convolutional Neural Networks for Classifying Lymphoma Using Histopathological Images**," *International Journal of Computer Applications*, vol. 182, no. 34, pp. 10–17, 2023.

[6] S. Pendhari and A. Thakare, "**Benchmarking Deep Learning Models for Automated MRI-based Brain Tumor Detection: CNN, VGG16, VGG19, ResNet-50, MobileNet, and InceptionV3**," *International Journal of Computer Applications*, vol. 186, no. 50, pp. 34–42, 2024.

[7] M. Sharif, J. Qadir, F. U. Samad, and S. Kadry, "**A hybrid deep CNN model for brain tumor image multi-classification**," *BMC Medical Imaging*, vol. 24, no. 1, pp. 1–16, 2024.

[8] R. Gupta, P. K. Singh, and R. K. Sahu, "**Revolutionizing Oral Cancer Detection: Aquila and Gorilla Algorithms Optimized Transfer Learning-Based CNNs (VGG16, MobileNet, DenseNet201)**," *Photonics*, vol. 8, no. 6, Art. no. 499, 2024.

[9] J. Lee and S. Kim, "**Histopathological Image Analysis for Oral Squamous Cell Carcinoma (OSCC) using MobileNetV3 and DenseNet**," *International Journal of Computer Assisted Radiology and Surgery*, vol. 19, no. 2, pp. 245–253, 2023.

[10] S. S. Liang, H. Wei, and M. Tan, "**A Multi-Modal Multi-Cancer Image Classification System Using Ensemble Deep Neural Networks (Leukemia, Lymphoma, Oral)**," *Applied Intelligence*, vol. 54, no. 3, pp. 2890–2907, 2024.

[11] A. Agarwal and R. M. Patel, "**Transfer Learning with Ensembles of Deep Neural Networks for Skin Cancer Detection in Imbalanced Datasets**," *arXiv preprint*, arXiv:2103.12068, 2021.