

Article

Not peer-reviewed version

Browser-Based Multi-Cancer Classification Framework Using Depthwise Separable Convolutions for Precision Diagnostics

[Divine Sebukpor](#), [Ikenna Odezuligbo](#)^{*}, Maimuna Nagey, Michael Chukwuka, Oluwamayowa Akinsuyi

Posted Date: 21 October 2025

doi: 10.20944/preprints202510.1612.v1

Keywords: deep learning; multi-cancer detection; medical imaging; tumor identification; explainable artificial intelligence; clinic; algorithmic transparency; radiomics



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Browser-Based Multi-Cancer Classification Framework Using Depthwise Separable Convolutions for Precision Diagnostics

Divine Sebukpor ¹, Ikenna Odezuligbo ^{2*}, Maimuna Nagey ³, Michael Chukwuka ⁴ and Oluwamayowa Akinsuyi ⁵

¹ Department of Computer Science, Chandigarh University, Punjab, India

² Department of Physics, Creighton University, Omaha, 68131, Nebraska, USA

³ Department of Medical Physics, Creighton University, Omaha, 68131, Nebraska, USA

⁴ Department of Physics and Astronomy, University of Kansas, Lawrence, 66045, Kansas, USA

⁵ Department of Microbiology and Cell Science, University of Florida, Gainesville, 32611, USA

* Correspondence: odezuligboe@gmail.com; Tel.: +14027088136

Abstract

Cancer is a major global health burden where early detection is paramount. While deep learning offers powerful diagnostic potential, its clinical adoption is often hindered by high computational costs, infrastructure demands, and data privacy concerns, particularly in low-resource settings. This study addresses these challenges by developing a highly accurate and universally accessible multi-cancer classification framework. We fine-tuned the Xception architecture on a comprehensive dataset of over 130,000 medical images across 26 cancer types and deployed it for client-side inference in a web browser using TensorFlow.js. The model's performance was benchmarked against Visual Geometry Group-16 (VGG16) and Residual Network-50 (ResNet50), with interpretability assessed using Grad-CAM. Our framework achieved outstanding performance with a Top-1 accuracy of 99.73%, significantly outperforming VGG16 and ResNet50. The browser-based tool enabled real-time, privacy-preserving inference, and Grad-CAM visualizations confirmed that predictions were based on clinically relevant features. This work demonstrates a viable paradigm for bridging the gap between advanced Artificial Intelligence (AI) and global health equity, offering a scalable, cost-effective, and private solution to democratize access to state-of-the-art cancer diagnostics.

Keywords: deep learning; multi-cancer detection; medical imaging; tumor identification; explainable artificial intelligence; clinic; algorithmic transparency; radiomics

1. Introduction

Cancer represents one of the most significant and formidable public health challenges of the 21st century. As a leading cause of morbidity and mortality worldwide, it is responsible for an estimated 10 million deaths annually, a figure that continues to rise amid global demographic shifts[1]. The clinical and economic burden of cancer is immense, placing extraordinary strain on healthcare systems, economies, and societies across the globe. However, this burden is not uniform across different cancer types, as illustrated in Figure 1. Diagnosing cancer at its early stages dramatically increases the likelihood of successful treatment and long-term survival. Conversely, delays in diagnosis allow for disease progression, leading to more complex and aggressive treatment regimens, higher healthcare costs, and significantly lower survival rates. This imperative for early detection has catalyzed extensive research into novel diagnostic technologies and screening programs. However, significant barriers remain, especially in low- and middle-income countries (LMICs), where limited access to specialist pathologists and advanced equipment creates profound disparities in patient care [2,3].

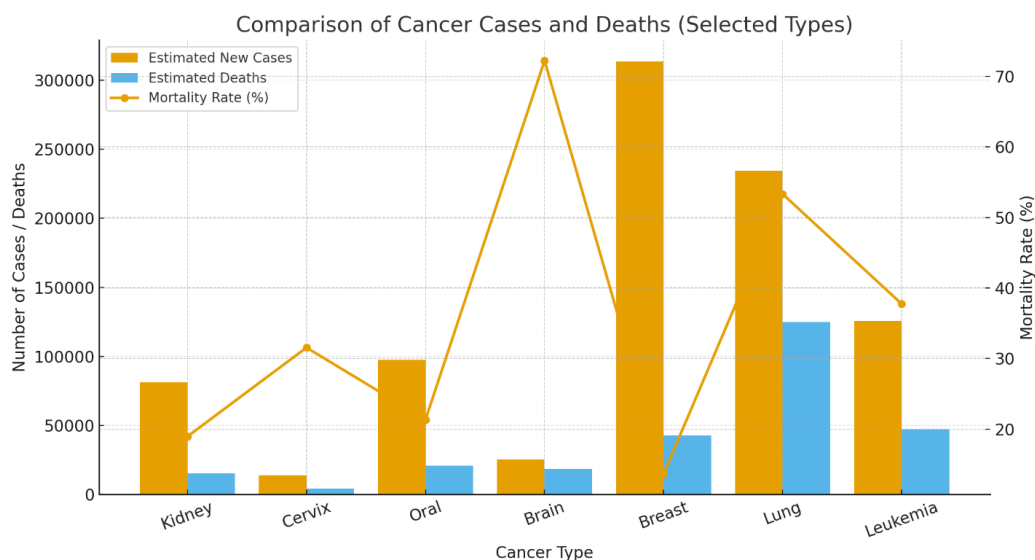


Figure 1. Comparison of Estimated New Cancer Cases, Deaths, and Mortality Rates for Selected Cancers. The bar chart displays the number of estimated new cases and deaths for six major cancer types, while the red line plot shows the corresponding mortality rate (%).

Medical image analysis has been transformed in the past decade by the rapid progress of deep learning, particularly through Convolutional Neural Networks (CNNs). CNNs have demonstrated a remarkable capacity to learn intricate patterns from visual data[4], offering the potential to augment the diagnostic process with quantitative, reproducible, and rapid analysis that can complement the assessments of human experts[5]. Early applications in oncology were often highly specialized tools trained for a single task, such as binary classification (e.g., malignant vs. benign)[6]. Recognizing this limitation, the research community has increasingly shifted toward developing comprehensive, multi-cancer classification systems designed to simultaneously differentiate among a wide spectrum of cancer types from a single analysis, representing a significant step toward a truly versatile clinical support tool[7].

Recent efforts in the field have leveraged specialized CNN structures and advanced techniques to improve performance. The Multi-Scale Feature Fusion Deep Convolutional Neural Network on Cancerous Tumor Detection and Classification (MFFDCNN-CTDC) model by Prakash et al. (2025)[8] exemplifies this trend by combining ResNet50 and EfficientNet backbones for comprehensive feature extraction, CAEs for hierarchical classification, and advanced parameter tuning through hybrid optimization techniques. Validated on large-scale melanoma image datasets, the MFFDCNN-CTDC achieved state-of-the-art test accuracies of 98.78% and 99.02%. Hybrid CNN ensemble models, integrating vision transformers and InceptionV3, have been proposed by Habeeba and Mahabubullah (2025)[9] for multi-cancer diagnosis across histopathological and clinical imaging datasets (e.g., brain, oral, lung), attaining improved accuracy and scalability, and deployed with web interfaces for fast clinical screening. Nasir et al.(2025)[10] performed extensive studies and validated CNN's effectiveness not just in skin cancer, but also in breast, lung, prostate, and colorectal cancers. In breast cancer detection, VGG16, ResNet, EfficientNet, and ensemble CNNs achieved AUC scores up to 0.99 and classification accuracies between 89% and 98%, outperforming traditional methods and supporting early, automated diagnosis. Furthermore, transfer learning with pre-trained networks enabled CNNs to generalize across small and imbalanced cancer datasets, boosting diagnosis for rare or underrepresented cancers. Jian et al. (2025)[11] Lung cancer detection frameworks using CNNs (and hybrid CNN–Artificial Neural Network, CNN–Long Short-Term Memory CNNsmodels) delivered diagnostic accuracies up to 99.54% on multiclass chest Computed Tomography (CT) images, with deep networks successfully integrating radiological feature extraction and clinical context. In skin and prostate cancer, CNNs paired with data augmentation and feature selection demonstrated high classification accuracy, further validating CNNs' broad

applicability in multi-cancer detection. Nagvi et al. (2023)[12] reported the viability of embedding deep learning models into web applications, allowing instant inference from uploaded medical images while ensuring patient privacy. Mobile platforms, leveraging models like MobileNetV2 and DenseNet variants, have enabled accurate classification at the point of care, facilitating rapid screening and telemedicine support in low-resource settings.

Despite the proven accuracy of these models, an important, though often overlooked, challenge in the translation of AI from research to clinical practice is the "last mile" problem of deployment. State-of-the-art deep learning models are computationally demanding, typically requiring specialized hardware and complex server-side. This creates a significant barrier to adoption in under-resourced clinics and hospitals in LMICs that lack the capital and expertise to maintain such systems. Furthermore, traditional cloud-based deployment models raise substantial concerns regarding data privacy, security, and compliance with stringent healthcare regulations, as they require uploading sensitive patient data to a remote server. In response to this deployment gap, a growing body of research has focused on architectural efficiency. Models incorporating depthwise separable convolutions and custom lightweight networks have achieved high accuracy while significantly reducing computational burden. This efficiency has enabled a new deployment paradigm: running inference directly within a web browser using frameworks like TensorFlow.js. This client-side approach allows for instant analysis while ensuring patient privacy, making advanced AI tools accessible even in low-resource settings.

The main aim of this work is to bridge the gap between high-accuracy models and accessible deployment by developing and validating a multi-cancer classification framework. The primary objectives are: (1) to develop and evaluate a deep learning framework using the Xception architecture for classifying 26 distinct cancer types from a composite dataset of over 130,000 medical images; (2) to conduct a comprehensive comparative analysis against foundational CNNs (VGG16, ResNet50) to assess accuracy and computational efficiency; and (3) to design and implement a novel deployment pipeline using TensorFlow.js that enables real-time, private, client-side inference directly within a web browser. The main conclusion of this study is that the combination of a computationally efficient architecture and a browser-based deployment strategy provides a viable and scalable solution to democratize access to advanced diagnostic AI, addressing critical barriers in global health.

2. Materials and Methods

This study adopts a systematic framework to develop and evaluate a multi-cancer classification system using deep learning. The methodology focuses on fine-tuning the Xception architecture and is outlined across several distinct stages to ensure clarity and reproducibility. These stages include dataset curation and preprocessing, a comparative analysis of model architectures, a detailed experimental protocol for training and evaluation, and a novel client-side deployment strategy.

2.1. Dataset Curation and Characteristics

The images for this research were collated from eight separate, publicly available datasets, each focusing on a specific category of cancer: Acute Lymphoblastic Leukemia (ALL)[13], Malignant Lymphoma[14], Brain Tumor MRI[15], Lung and Colon Cancer (LC25000) [16], Kidney CT (Normal-Cyst-Tumor-Stone)[17], Cervical Cancer (SIPaKMeD)[18], Oral Squamous Cell Carcinoma (OSCC)[19], Breast Cancer (BreakHis)[20].

2.2. Image Preprocessing and Augmentation Pipeline

A standardized preprocessing and augmentation pipeline was applied to all images to ensure consistency and enhance model robustness.

Preprocessing:

Before being fed into the neural networks, every image in the dataset underwent two preprocessing steps. First, all images were resized to a uniform dimension of 224×224 pixels. This step is essential to match the required input size of the pre-trained architectures (VGG16, ResNet50, and Xception), which were originally trained on ImageNet using this dimension. Second, the pixel values of each image, which are typically encoded as integers from 0 to 255, were normalized to a floating-point range by dividing each pixel value by 255.

Augmentation:

To artificially expand the diversity of the training dataset and improve the model's ability to generalize to unseen data, a comprehensive suite of online data augmentation techniques was employed during the training phase. Online augmentation applies random transformations to each batch of images as it is fed to the model, ensuring that the network rarely sees the exact same image twice. This process helps to prevent overfitting by teaching the model to be invariant to minor variations in position, orientation, and lighting that are commonly encountered in real-world medical imaging. These techniques were chosen to simulate realistic variations while preserving the core diagnostic features of the pathologies.

A key characteristic of the curated dataset after preprocessing is its balanced class distribution, which was intentionally designed to mitigate the risks of model bias that often arise from imbalanced data. The dataset was partitioned into three distinct subsets: a training set, a validation set, and a test set. The training set was constructed to contain exactly 4,000 images for each of the 26 classes. The validation and test sets were each designed to contain 500 images per class, with minor exceptions for two classes that had 501 images in the test set. The detailed distribution of images for each class across the training, validation, and test sets is presented in Table 1. All data was sourced from publicly accessible repositories, and as such, patient-identifying information had been removed at the source, adhering to ethical standards for research involving medical data.

Table 1. Detailed Distribution of the 26 Cancer Classes Across Training, Validation, and Test Sets.

Full Cancer Type Name	Train Images	Val Images	Test Images	Total Images
Acute Lymphoblastic Leukemia Benign	4000	500	500	5000
Acute Lymphoblastic Leukemia Early	4000	500	500	5000
Acute Lymphoblastic Leukemia Pre	4000	500	500	5000
Acute Lymphoblastic Leukemia Pro	4000	500	500	5000
Brain Glioma	4000	500	500	5000
Brain Meningioma	4000	500	500	5000
Brain Tumor	4000	500	500	5000
Breast Benign	4000	500	500	5000
Breast Malignant	4000	500	500	5000
Cervix Dyskeratotic	4000	500	500	5000

Cervix Koilocytotic	4000	500	500	5000
Cervix Metaplastic	4000	500	500	5000
Cervix Parabasal	4000	500	500	5000
Cervix Superficial Intermediate	4000	500	500	5000
Colon Adenocarcinoma	4000	500	500	5000
Colon Benign Tissue	4000	500	500	5000
Kidney Normal	4000	500	500	5000
Kidney Tumor	4000	500	500	5000
Lung Adenocarcinoma	4000	500	500	5000
Lung Benign Tissue	4000	500	500	5000
Lung Squamous Cell Carcinoma	4000	500	500	5000
Chronic Lymphocytic Leukemia	4000	500	500	5000
Follicular Lymphoma	4000	500	500	5000
Mantle Cell Lymphoma	4000	500	500	5000
Oral Normal	4000	500	501	5001
Oral Squamous Cell Carcinoma	4000	500	501	5001

2.3. Model Architectures

To address the multi-cancer classification problem, we adopted a **transfer learning framework** built upon three state-of-the-art convolutional neural network (CNN) architectures-**Xception**, **VGG16**, and **ResNet50**, each chosen for its unique balance between representational depth, computational efficiency, and architectural innovation.

2.3.1. Primary Architecture: Xception

The primary model investigated in this study is based on the Xception architecture. The core innovation of Xception is its extensive use of depthwise separable convolutions, which serve as a replacement for the standard convolution layers found in earlier architectures. A depthwise separable convolution factorizes a standard convolution into two distinct operations: a depthwise convolution that applies a single spatial filter to each input channel independently, and a pointwise convolution (1×1) that then computes a linear combination of the outputs of the depthwise convolution across channels [21].

Structurally, Xception comprises **three main flows**:

- **Entry Flow**, which extracts low-level features through depthwise convolutions and pooling;
- **Middle Flow**, containing multiple identical modules that refine hierarchical representations; and
- **Exit Flow**, where high-level features are aggregated and projected through global average pooling and dense layers.

The forward computation for a single Xception block can be formalized as:

$$\text{Output} = X + \sum_i \beta_i \cdot \text{BatchNorm}(W_i * X, \gamma_i), \quad (1)$$

where X is the input tensor, W_i are depthwise kernels, β_i and γ_i are the scaling and normalization parameters, respectively.

2.3.2. Comparative Architecture: VGG16 and ResNet50

For benchmarking, **VGG16** and **ResNet50** were implemented to evaluate the relative performance of older yet widely benchmarked CNN designs.

VGG16 is a sequential deep CNN consisting of 13 convolutional and 3 fully connected layers arranged in five blocks, each using small 3×3 filters and followed by 2×2 max-pooling operations. The generalized convolutional transformation in VGG16 can be expressed as:

$$\text{Output} = \text{Activation}(\text{Convolution}(W * X + b)), \quad (2)$$

where W represents the convolutional filter weights, b the bias term, and $\text{Activation}()$ is typically the ReLU nonlinearity. The simplicity and uniformity of its design are its main strengths, but its large number of parameters (~138 million) makes it computationally intensive[22]

ResNet50 introduces residual learning to overcome the vanishing gradient problem inherent in very deep networks. Each residual block learns a residual mapping $F(X)$ relative to its input, enabling direct gradient flow through skip connections[23]. The output of a residual block can be defined as:

$$\text{Output} = \text{Activation} \left(\text{BatchNorm}(W_2 * \text{Activation}(\text{BatchNorm}(W_1 * X + b_1)) + b_2) \right) + X. \quad (3)$$

This formulation allows the network to learn residual functions rather than full transformations, improving convergence and generalization. The bottleneck block structure ($1 \times 1 \rightarrow 3 \times 3 \rightarrow 1 \times 1$) further enhances efficiency by reducing the number of feature maps before costly spatial convolutions.

2.4. Experimental Protocol

To ensure reproducibility, a standardized experimental protocol was followed for training and evaluating all models. The end-to-end workflow is summarized in Figure 2.

Hardware and Software Environment:

All training and evaluation experiments were conducted on a high-performance computing system equipped with an NVIDIA A100 Graphics Processing Units (GPU), which is designed to accelerate deep learning workloads. The software environment was built using Python, with the deep learning models implemented, trained, and evaluated using the TensorFlow framework and its high-level Keras API.

Training, Optimization, and Regularization:

The models were trained using the Adam optimizer, a widely used adaptive learning rate optimization algorithm, with an initial learning rate set to 1×10^{-4} . The training process was configured with a batch size of 32 and ran for a total of 21 epochs. The loss function employed was categorical cross-entropy, which is the standard choice for multi-class classification problems, as it measures the divergence between the predicted probability distribution and the true one-hot encoded label.

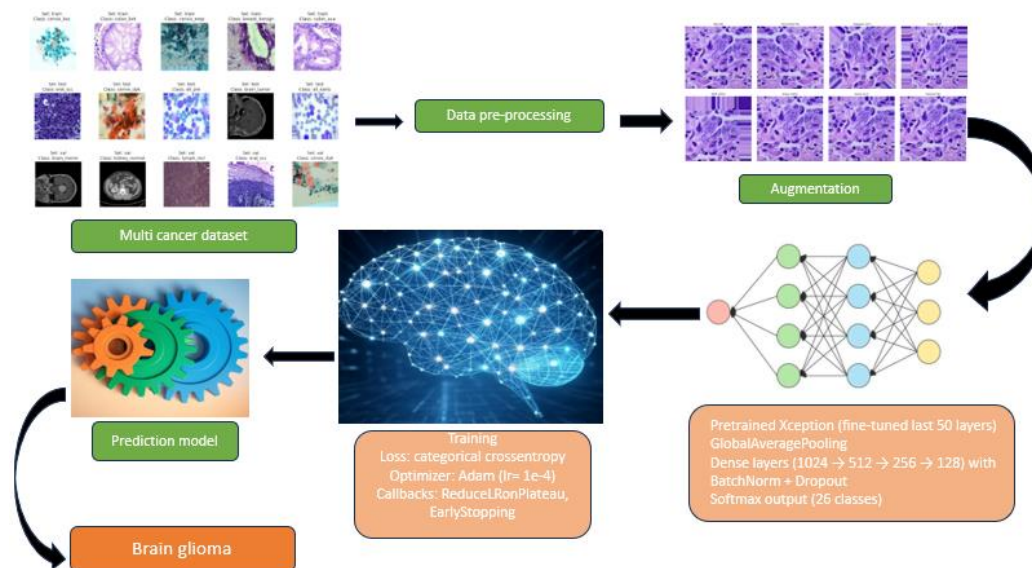


Figure 2. End-to-end training pipeline for multi-cancer classification.

To further optimize the training process and enhance model generalization, two key Keras callbacks were implemented:

EarlyStopping: This callback was configured to monitor the validation loss. If the validation loss did not show improvement for a "patience" of 10 consecutive epochs, the training process would be automatically halted. This prevents the model from overfitting to the training data and saves computational resources by stopping training once performance on unseen data plateaus[24].

ReduceLRonPlateau: This callback also monitored the validation loss. If the loss failed to improve for 5 consecutive epochs, the learning rate would be reduced by a factor of 0.2. This strategy, known as learning rate scheduling, allows the model to take large steps in the parameter space early in training and smaller, more refined steps as it approaches a minimum, often leading to better final performance[25].

Evaluation Metrics:

A comprehensive suite of metrics was used to evaluate the performance of the trained models on the held-out test set [26,27]. These included:

Top-1 Accuracy: It measures the proportion of predictions where the model's single highest-probability guess is the correct one. Given a dataset of N examples, each with a true label and a model's prediction, the top-1 accuracy is calculated as:

$$\text{Top - 1 Accuracy} = \frac{\text{Number of correct top-1 predictions}}{\text{Total number of predictions}}. \quad (4)$$

Top-5 Accuracy: considers a prediction correct if the true class is among the top five most probable classes predicted by the model.

$$\text{Top - 5 Accuracy} = \frac{1}{N} \sum_{i=1}^N I(y_i \in C_{i,5}), \quad (5)$$

where N be the total number of test samples, y_i be the true class for the sample i , $C_{i,k}$ be the set of the top k predicted classes for sample i , and I be an indicator function that returns 1 if its argument is true and 0 otherwise.

Macro-Averaged Precision, Recall, and F1-Score: To assess performance across all 26 classes in a balanced manner, the precision, recall, and F1-score were calculated for each class individually and then averaged (macro-average). This approach gives equal weight to each class, regardless of its size, providing a robust measure of overall performance on a multi-class problem.

• Classes $C = \{1, \dots, 26\}$

• For class c : TP_c, FP_c, FN_c from one-vs-rest confusion counts

• Per-class precision/recall/F1:

$$\text{Prec}_c = \frac{TP_c}{TP_c + FP_c}, \text{Rec}_c = \frac{TP_c}{TP_c + FN_c}, \text{F1}_c = \frac{2\text{Prec}_c\text{Rec}_c}{\text{Rec}_c + \text{Prec}_c}. \quad (6)$$

Macro averages (unweighted mean over classes)

$$\text{Macro - Prec} = \frac{1}{|C|} \sum_{c \in C} \text{Prec}_c, \text{Macro - Rec} = \frac{1}{|C|} \sum_{c \in C} \text{Rec}_c,$$

$$\text{Macro - F1} = \frac{1}{|C|} \sum_{c \in C} \text{F1}_c. \quad (7)$$

2.5. Client-Side Deployment via TensorFlow.js

To enable privacy-preserving, low-latency use without dedicated servers, we deployed the trained network for **in-browser inference** using TensorFlow.js. The Keras/TensorFlow model was exported to the TensorFlow **SavedModel** format and converted to a TensorFlow.js **graph model** comprising a JSON graph descriptor and sharded binary weights. The resulting assets were served from a static host and loaded at runtime by the client application (HTML/JavaScript).

On the client, images are read from the browser file API, converted to tensors, **resized** to the network input resolution, and **normalized** with the same transformation as training. Inference executes entirely on the user's device (WebGL/WebGPU acceleration when available). The interface reports the **Top-1** prediction, **top-k** probabilities, and optionally exports a structured PDF summary for record-keeping. Because images never leave the device, this architecture minimizes data transfer, mitigates regulatory risk, and supports **offline** use after the initial model fetch.

To ensure parity across runtimes, we (i) duplicated the preprocessing pipeline in JavaScript, (ii) validated numerics against Python exports on a held-out set, and (iii) version-locked converter/runtime packages.

2.6. Explainability and Visualization

Interpretable artificial intelligence is essential for the clinical integration of deep learning systems. In this study, we adopted **Gradient-weighted Class Activation Mapping (Grad-CAM)** to visualize how the network arrived at its classification decisions. Grad-CAM generates a **heatmap** that identifies the image regions contributing most strongly to a model's prediction by leveraging the gradient information from the final convolutional layer [28].

Mathematically, the Grad-CAM map for a target class c can be expressed as:

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right), \quad (8)$$

where A^k is the k -th feature map of the final convolutional layer, and α_k^c denotes its relative importance, computed via global average pooling of the gradients of the class score with respect to A^k . The **ReLU** operation retains only the features exerting a positive influence on the target class.

The resulting map is resized to match the input image and overlaid for visual interpretation. This enables radiologists to assess whether the model's highlighted regions correspond to diagnostically relevant structures or lesions, strengthening confidence in its clinical applicability.

3. Results

This section presents a comprehensive evaluation of the proposed multi-cancer classifier. We first report aggregate accuracy metrics (top-1/top-5) and macro-averaged precision, recall, and F1 across all 26 classes using the held-out test split, followed by class-wise results and confusion matrix analysis to expose residual failure modes and visually similar differentials. We then benchmark Xception against VGG16 and ResNet50 under matched training protocols. Finally, we report inference throughput in TensorFlow.js and summarize the explainability outcomes of Grad-CAM overlays.

3.1. Quantitative Performance of the Fine-Tuned Xception Model

The primary model, based on the fine-tuned Xception architecture, demonstrated outstanding performance on the held-out test set. The model achieved a **Top-1 accuracy of 99.73%**, indicating that its highest-confidence prediction was correct for the vast majority of cases. Even more impressively, the model achieved a **Top-5 accuracy of 100.00%**, signifying that for every single image in the test set, the true class was included within the model's top five predictions. Aggregate results are summarized in **Figure 3**.

To further assess the model's robustness and balance across all 26 classes, macro-averaged precision, recall, and F1-score were calculated. The model achieved a **macro-averaged precision of 1.00**, a **macro-averaged recall of 1.00**, and a **macro-averaged F1-score of 1.00**. These perfect scores indicate that the model not only avoided false positives (high precision) but also successfully identified all true positive cases (high recall) for every class, resulting in a perfectly balanced and highly reliable classification performance across the entire dataset.

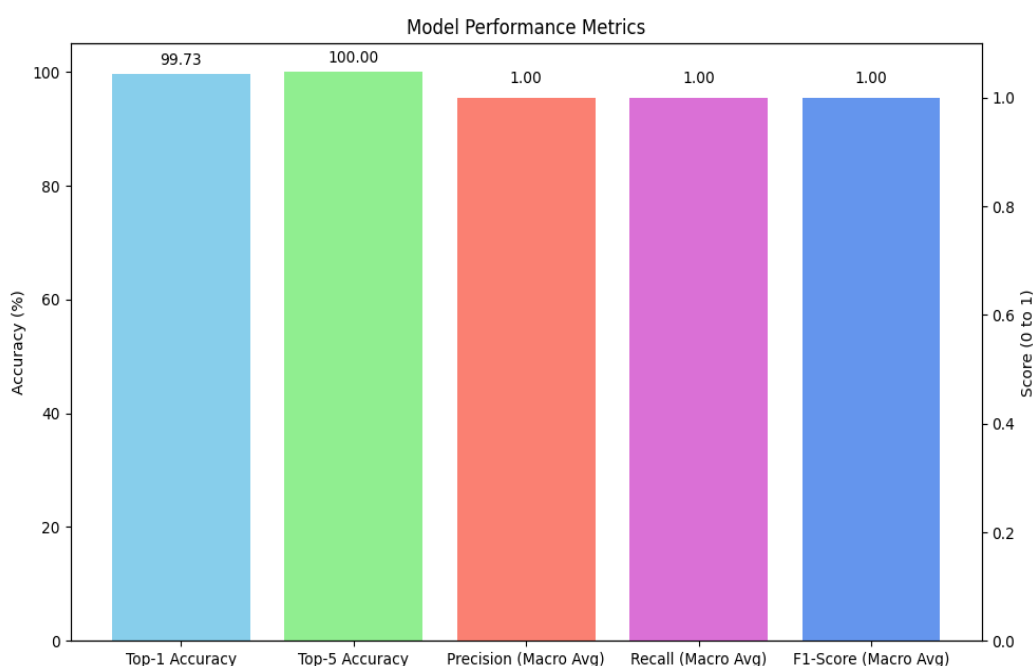


Figure 3. Aggregate performance of the Xception classifier on the held-out test set.

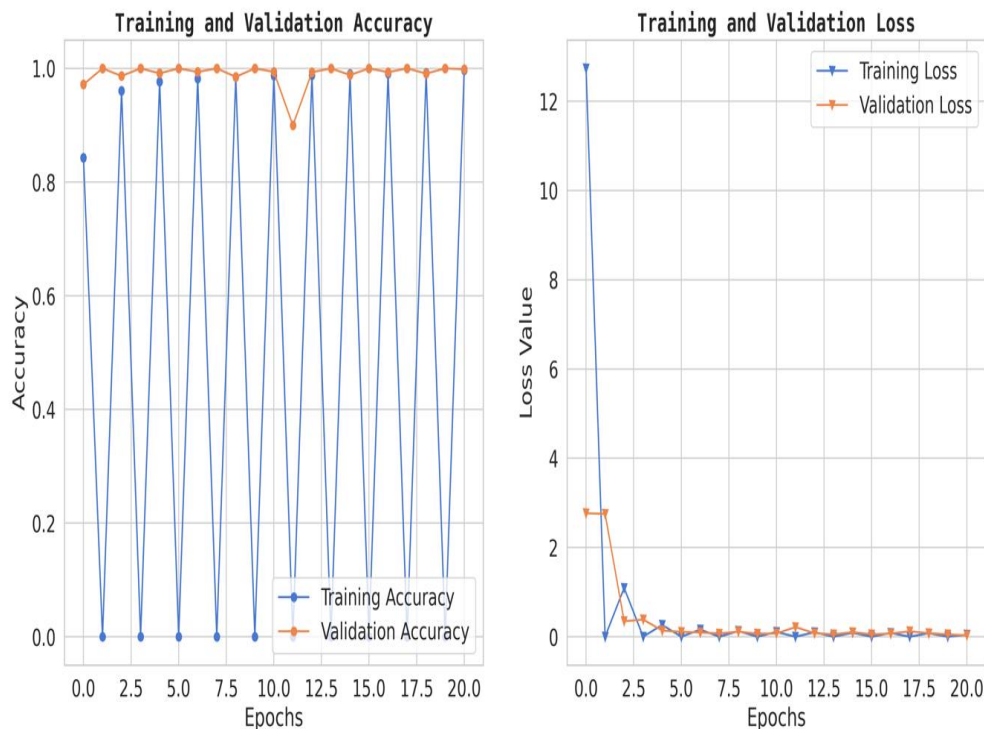


Figure 4. Training dynamics of the Xception model over 21 epochs.

The training and validation curves, plotted over 21 epochs, provide insight into the model's learning dynamics. The training accuracy rapidly approached 100%, while the validation accuracy closely tracked it, reaching a high plateau and exhibiting minimal divergence. Similarly, the training loss decreased steadily and converged to a value near zero, while the validation loss remained low and stable throughout the training process. This behavior demonstrates that the model learned effectively from the training data without significant overfitting, successfully generalizing its learned features to the unseen validation data. The stability of the validation curves confirms the efficacy of the regularization techniques (Dropout, L2 regularization) and the data augmentation pipeline. Learning dynamics are shown in **Figure 4**.

3.2. Class-Specific Performance and Confusion Matrix Analysis

A detailed examination of the model's class-specific performance was conducted by analyzing the confusion matrix generated from the test set predictions, as shown in Figure 5. The matrix provides a granular view of the model's accuracy, with the diagonal elements representing correct classifications and off-diagonal elements representing misclassifications.

The confusion matrix reveals a strong performance, with nearly all predictions falling along the main diagonal. This visualizes the model's exceptional ability to accurately distinguish between the 26 different cancer types. However, a small number of off-diagonal entries highlight the few instances of confusion. Specifically, the model exhibited minor

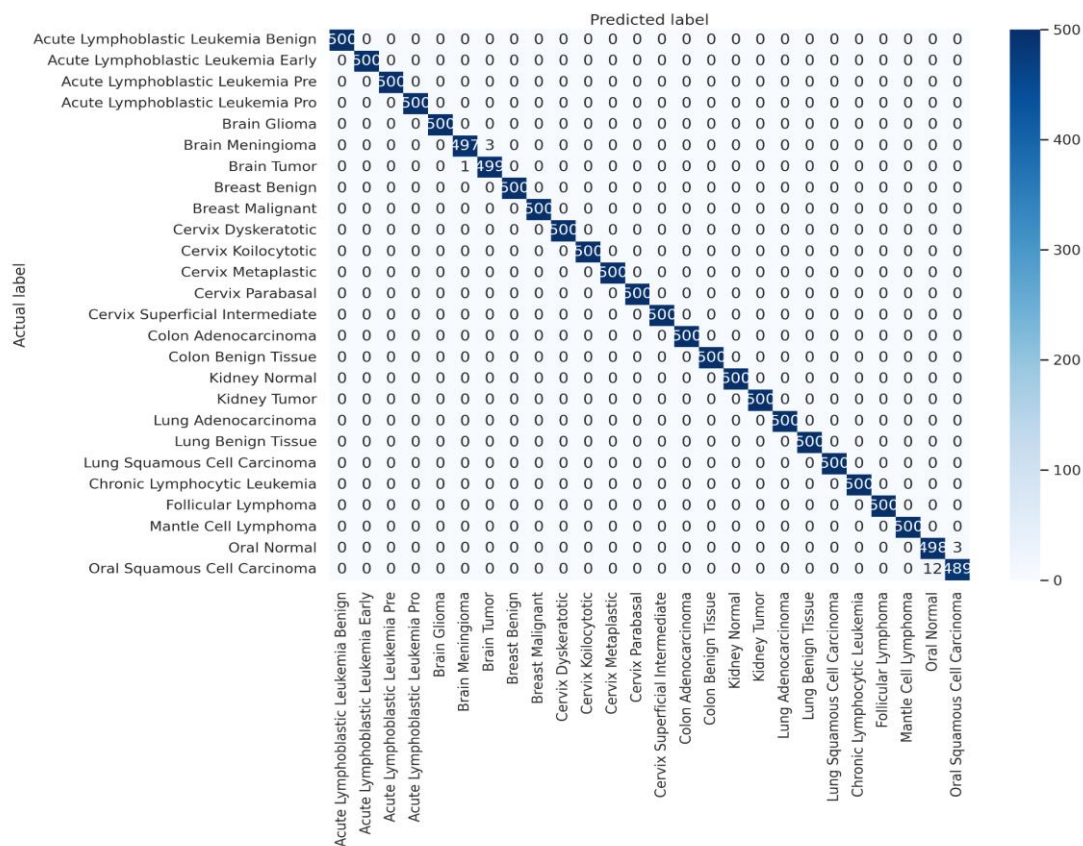


Figure 5. Confusion matrix for 26-class multi-cancer classification on the test set.

confusion between classes with high morphological similarity or those originating from the same anatomical location. For example, there were rare misclassifications between 'Brain Glioma' and 'Brain Meningioma', two distinct types of brain tumors that can present with overlapping visual features in medical images. Similarly, slight confusion was observed between 'Oral Squamous Cell Carcinoma' and 'Oral Normal' tissue. These isolated instances of error are instructive, as they pinpoint the most challenging diagnostic distinctions for the model and suggest that its residual weaknesses lie in differentiating pathologies with very subtle visual cues. Nevertheless, the extremely low rate of such misclassifications underscores the model's overall robustness and high degree of diagnostic precision.

3.3. Model Interpretability Using Grad-CAM

Our analysis of the Grad-CAM visualizations confirms that the model bases its predictions on clinically relevant features. As illustrated in the representative cases, the heatmaps consistently localized on regions with clear pathological characteristics. For the lung adenocarcinoma classification (Figure 7), the activation map precisely highlights a cluster of atypical cells with large, irregular nuclei. Similarly, in the breast malignancy case (Figure 6), the model's attention is concentrated on an area of high cellular density and disorganized tissue architecture, consistent with invasive carcinoma. This strong alignment between the model's focus and known histopathological indicators provides compelling evidence that the model has learned to identify legitimate disease patterns rather than relying on spurious artifacts.

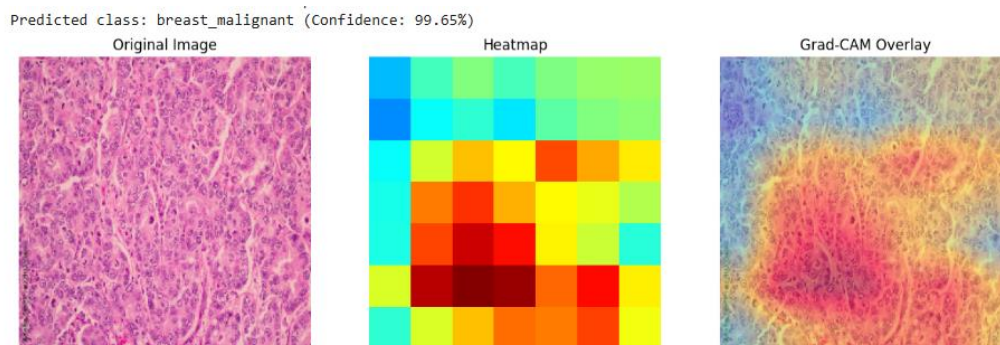


Figure 6. Grad-CAM visualization for a breast malignant classification (99.65% confidence). The heatmap and overlay show the model's attention is concentrated on sheets of infiltrating malignant cells, a hallmark of invasive breast cancer.

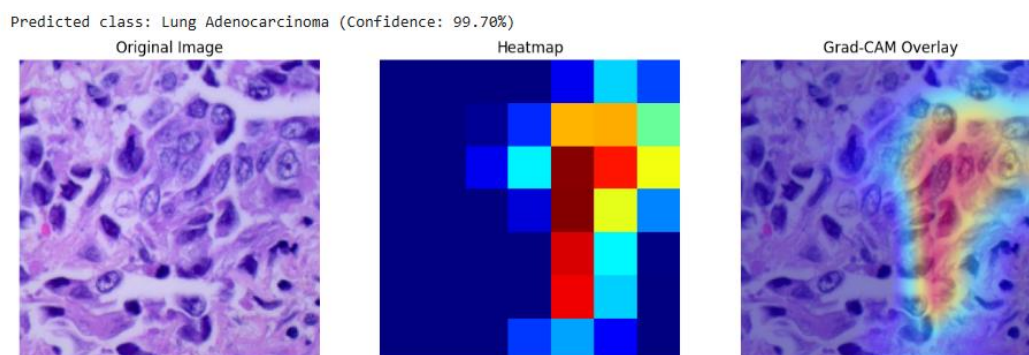


Figure 7. Grad-CAM visualization for a lung adenocarcinoma classification (99.70% confidence). The overlay (right) confirms the model's focus on a region of atypical glandular cells, which are critical for diagnosis.

3.4. Comparative Analysis of Deep Learning Architectures

All three models were trained and evaluated under identical conditions, including the same dataset, preprocessing pipeline, fine-tuning strategy, and training hyperparameters, to ensure a fair and direct comparison.

As shown in Figure 8, the Xception model significantly outperformed both VGG16 and ResNet50 in Top-1 accuracy, achieving a score over 5 percentage points higher than its closest competitor, ResNet50. While VGG16 and ResNet50 delivered respectable results, with Top-1 accuracies of 93.50% and 94.20% respectively, they did not approach the near-perfect performance of Xception. Furthermore, only the Xception model achieved a perfect Top-5 accuracy of 100.00%, highlighting its unique reliability in positioning the correct diagnosis within its top predictions.

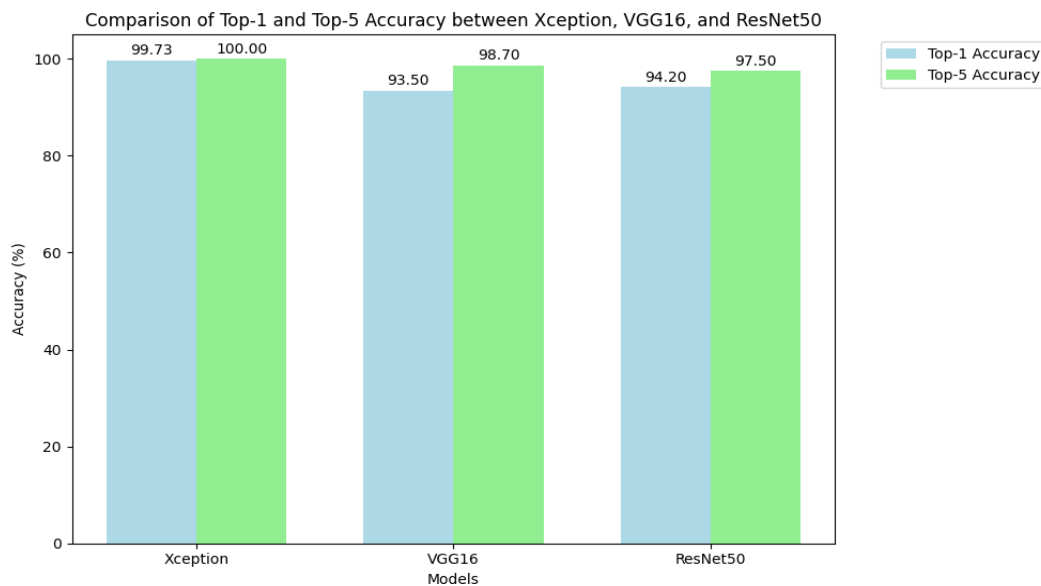


Figure 8. The above graph shows the comparison in the top-1 and top-5 accuracy of the Xception, VGG16 and ResNet50 on multi-cancer dataset used for training and testing.

In addition to classification accuracy, the models were compared based on their architectural complexity and computational cost. This analysis, presented in Table 2, reveals that Xception's superior performance is achieved with remarkable efficiency.

Table 2. Comparison of Model Complexity and Computational Cost (Parameters, Model Size, Relative Training Time).

Model	Parameters (Millions)	Model Size (MB)	Total Training Time (Hours)
Xception	~23.9	~90	10
VGG16	~138.4	~528	30
ResNet50	~25.6	~98	16

Table 2 shows the architectural efficiency of Xception. It contains a similar number of parameters to ResNet50 and is over six times smaller than VGG16 in terms of parameter count.⁷ This lean architecture translates directly into a smaller memory footprint and faster training times. While VGG16 is known to be "painfully slow to train" due to its massive fully connected layers⁸, the Xception model completed its training in 10 hours on an NVIDIA A100 GPU. This combination of superior accuracy and high computational efficiency makes Xception the unequivocally best performing and most practical architecture for the task among the models evaluated.

3.5. Performance of the Deployed Web-Based Tool

The final stage of the results validation involved confirming the successful deployment and functionality of the trained Xception model via the TensorFlow.js pipeline. The conversion from the Keras SavedModel format to the TensorFlow.js graph model format was executed without errors. The resulting web application was tested across multiple standard web browsers (e.g., Google Chrome, Mozilla Firefox).

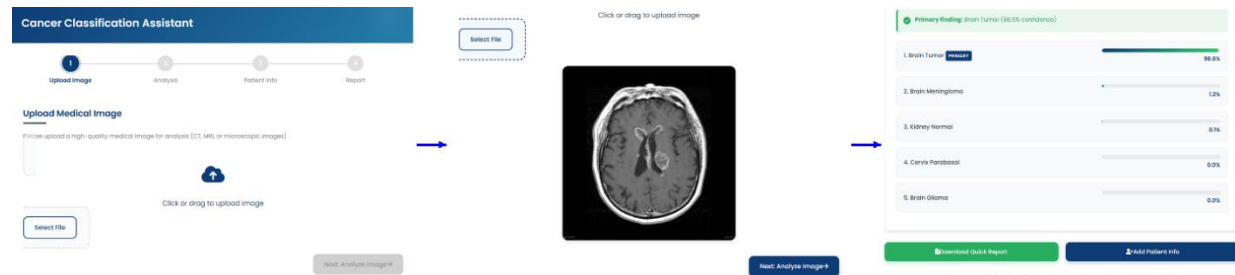


Figure 9. Workflow of the Cancer Classification Assistant. The interface demonstrates the sequential pipeline from (A) **Image Upload**, where a CT/MRI/microscopic image is provided, to (B) **Automated Image Analysis**, which generates diagnostic predictions, and finally (C) **Classification Report**, presenting the primary finding with model confidence.

The deployed tool successfully loaded the model.json and associated weight files, initializing the model within the browser's runtime environment. Upon uploading test images from the held-out dataset, the application provided near-instantaneous classification results directly on the user interface, as shown in Figure 9. The entire process, from image selection to the display of the predicted class and confidence scores, occurred entirely on the client-side, with no data transmission to any backend server, thus validating the privacy-preserving design. The functionality of the user interface, including the image upload mechanism, real-time prediction display, and the generation of a downloadable PDF report, performed as designed. This successful deployment confirms the viability of using TensorFlow.js to create accessible, real-time, and secure diagnostic support tools based on complex deep learning models.

4. Discussion

This section interprets the empirical findings and situates them in the context of clinical use and deployment.

4.1. Interpretation of Findings: The Architectural Advantage of Xception

The experimental results present the superiority of the Xception architecture in this multi-cancer classification task. The model not only achieved higher accuracy than both VGG16 and ResNet50 but did so with comparable or superior computational efficiency. This outcome can be attributed to a fundamental alignment between Xception's architectural design and the specific nature of the data being analyzed. Histopathological and cytological images, which form the bulk of the dataset, are fundamentally characterized by intricate textures, fine-grained patterns, and subtle morphological variations, rather than the composition of distinct, large-scale objects found in general-purpose datasets like ImageNet.

Traditional convolutional architectures, such as VGG16 and ResNet50, employ standard convolution operations that simultaneously attempt to learn spatial hierarchies (e.g., identifying shapes and patterns) and cross-channel correlations (e.g., understanding how different learned features relate to one another). This conflation of tasks within a single operation, while effective for many problems, may not be the most efficient or effective way to model the nuanced features of medical images. Xception's core innovation, the depthwise separable convolution, explicitly

decouples these two learning processes. The depthwise step first learns spatial patterns within each feature map independently, effectively acting as a highly specialized texture and pattern detector. The subsequent pointwise step then learns how to best combine these detected patterns across all channels. This factorization allows the model to dedicate its parameters more efficiently to the task of learning the subtle textural variations that differentiate one cancer subtype from another. This architectural suitability for histopathology likely explains its superior feature extraction capabilities and, consequently, its higher classification accuracy. VGG16, with its massive parameter space, may struggle to learn these fine-grained features without overfitting, while ResNet50's architecture is primarily optimized for solving the degradation problem in very deep networks, not necessarily for maximizing feature efficiency in this specific context.

4.2. Model Interpretability and Clinical Trust

A significant barrier to the adoption of AI in clinical practice is the "black box" problem, where models provide highly accurate predictions without offering insight into their decision-making process. This study addresses this challenge through the use of Gradient-weighted Class Activation Mapping (Grad-CAM). As demonstrated in Section 3.3, the Grad-CAM visualizations confirm that the model's predictions are not arbitrary but are grounded in clinically relevant features. For instance, in cases of lung adenocarcinoma and breast malignancy, the model's attention was precisely localized on regions of atypical cells and disorganized tissue architecture.

This alignment between the model's focus and known histopathological indicators is critical for building clinical trust. It provides compelling evidence that the model has learned to identify legitimate disease patterns rather than relying on spurious artifacts or dataset-specific noise. By offering this layer of transparency, the framework moves beyond a simple prediction tool to become an interpretable diagnostic assistant, allowing clinicians to validate the AI's "reasoning" against their own expertise. This interpretability is essential for responsible clinical integration, as it empowers medical professionals to use the tool with greater confidence and understanding.

4.3. Clinical Significance and Potential Applications

In a real-world diagnostic setting, a pathologist's workflow rarely culminates in a single, instantaneous diagnosis. More commonly, especially in complex or ambiguous cases, the process involves formulating a differential diagnosis. The AI tool developed in this study is aligned with this workflow.

A Top-1 accuracy of 99.73%, still carries a minuscule but non-zero risk of providing an incorrect primary prediction. However, a Top-5 accuracy of 100% provides a guarantee: the correct diagnosis is always present within the top five suggestions generated by the model. This transforms the tool's role from that of a simple "classifier" to a "diagnostic assistant." It functions as an invaluable safety net, particularly for less experienced pathologists or in high-throughput laboratories where the risk of cognitive error or oversight is elevated. For instance, if a rare cancer presents atypical features resembling a more common condition, a human expert might anchor on the more common diagnosis. In such a scenario, the AI tool, having been trained on a vast and diverse dataset, would ensure that the rare possibility is still presented to the clinician for consideration. This capability to build a great differential diagnosis list has the potential to reduce diagnostic errors, accelerate the time to correct diagnosis, and ultimately improve patient care by ensuring all relevant diagnostic avenues are explored.

4.4. The Paradigm Shift of Browser-Based AI in Global Health

By performing all computations on the user's local machine, the model operates at the "edge." This has several transformative implications:

Zero Infrastructure Cost: There are no server costs for the provider and no need for specialized hardware at the clinic. Any standard computer with a web browser is sufficient to run the tool.

Absolute Data Privacy: Since the medical images are processed locally and never leave the user's device, patient privacy is guaranteed by design. This eliminates the regulatory and ethical quagmire of data transfer.

Universal Accessibility: The tool is accessible to anyone with an internet connection to download the web page, after which it can even function offline. This democratizes access to a state-of-the-art diagnostic tool, making it available to clinicians regardless of their geographic location or institutional resources.

This deployment strategy is a core contribution of the work, offering a viable and scalable blueprint for the equitable distribution of medical AI technologies and directly addressing the socioeconomic barriers that perpetuate global health disparities.

4.5. Limitations and Future Research Directions

While this study demonstrates significant technical and practical advancements, it is essential to acknowledge its limitations and outline avenues for future research to ensure a path toward responsible clinical integration.

First, the model was trained and validated on a curated, aggregated dataset composed of images from public repositories. Although large and diverse, this dataset may not fully represent the variability of "in-the-wild" clinical data, which can be affected by differences in patient demographics, imaging hardware, slide preparation techniques, and staining protocols across different institutions. Therefore, the model's generalizability to new, unseen clinical environments is not yet proven. A critical next step is to conduct extensive external validation using data from multiple, diverse clinical sites.

Second, the image preprocessing pipeline, while effective, was relatively simple, consisting of standard resizing and normalization. Future work could explore whether incorporating advanced preprocessing steps could further improve the model's performance, particularly for the few classes where minor confusion was observed.

Third, the current model is unimodal, relying exclusively on information derived from medical images. A more comprehensive diagnostic assessment could be achieved by integrating multi-modal data. Future iterations of the framework could be designed to incorporate complementary data sources such as genomic or proteomic data, histopathology reports, and patient clinical history to create a more holistic and accurate diagnostic profile.

Based on these limitations, future research will proceed along several key directions:

Prospective Clinical Trials: Conducting rigorous, multi-center prospective clinical trials is essential to validate the model's real-world efficacy and safety, which is a prerequisite for obtaining regulatory approvals for clinical use.

Dataset Expansion: The training dataset will be expanded to include rarer cancer types and a wider variety of imaging modalities (e.g., Magnetic Resonance Imaging, Positron Emission Tomography scans) to further improve the model's generalizability and broaden its clinical applicability.

Federated Learning: To enhance model robustness while preserving patient privacy across institutions, exploring federated learning is a promising direction. This approach would allow the model to be trained on decentralized data from multiple hospitals without the need to transfer any raw data to a central server.

5. Conclusions

This research successfully developed and validated a multi-cancer classification framework that demonstrates both diagnostic accuracy and an approach to accessible clinical deployment. By fine-tuning the computationally efficient Xception architecture on a comprehensive dataset of over 130,000 images across 26 cancer types, the model achieved a Top-1 accuracy of 99.73% and a Top-5 accuracy of 100.00%. The comparative analysis confirmed its superiority over established benchmark architectures like VGG16 and ResNet50, not only in accuracy but also in model efficiency. Crucially,

the model's reliability is further bolstered by interpretability analysis using Grad-CAM, which confirms that its predictions are based on clinically relevant histopathological features, building trust in its automated diagnostic process.

A key contribution of this work lies in the implementation of a deployment pipeline using TensorFlow.js that enables real-time, in-browser inference, effectively transforming the high-performance model into a universally accessible tool. The resulting tool provides not only rapid diagnostic support but also a degree of transparency, ensuring that sensitive medical data remains securely on the user's local device.

Looking forward, the framework provides a robust foundation for future enhancements, including the integration of multi-modal data sources, the exploration of more advanced interpretability techniques, and validation through large-scale clinical trials. The implication of this research is the demonstration of a viable pathway to bridge the gap between advanced AI and global health equity. By combining high diagnostic precision with a deployment model that is inherently private, scalable, and accessible, this work reinforces the transformative potential of AI-driven tools to reshape healthcare delivery, facilitate earlier and more accurate cancer detection, and ultimately improve patient outcomes for all, irrespective of geographic or economic barriers.

Author Contributions: Conceptualization, D.S. and I.O.; methodology, D.S.; writing—original draft preparation, I.O.; writing—review and editing, M.C.; visualization, M.N.; data curation and supervision, O.A.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used for this research can be assessed at <https://github.com/Sebukpor/multi-cancer-classification?tab=readme-ov-file>.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **2024**, *74*, 229–263, doi:10.3322/caac.21834.
2. Cotache-Condor, C.; Rice, H.E.; Schroeder, K.; Staton, C.; Majaliwa, E.; Tang, S.; Rice, H.E.; Smith, E.R. Delays in Cancer Care for Children in Low-Income and Middle-Income Countries: Development of a Composite Vulnerability Index. *Lancet Glob Health* **2023**, *11*, e505–e515, doi:10.1016/S2214-109X(23)00053-0.
3. Qu, L.G.; Brand, N.R.; Chao, A.; Ilbawi, A.M. Interventions Addressing Barriers to Delayed Cancer Diagnosis in Low- and Middle-Income Countries: A Systematic Review. *Oncologist* **2020**, *25*, e1382–e1395, doi:10.1634/theoncologist.2019-0804.
4. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* **2017**, *542*, 115–118, doi:10.1038/nature21056.
5. Tandon, R.; Agrawal, S.; Rathore, N.P.S.; Mishra, A.K.; Jain, S.K. A Systematic Review on Deep Learning-Based Automated Cancer Diagnosis Models. *J Cell Mol Med* **2024**, *28*.
6. Rhanoui, M.; Alaoui Belghiti, K.; Mikram, M. Multi-Task Deep Learning for Simultaneous Classification and Segmentation of Cancer Pathologies in Diverse Medical Imaging Modalities. *Onco* **2025**, *5*, 34, doi:10.3390/onco5030034.
7. Wang, C.W.; Khalil, M.A.; Firdi, N.P. A Survey on Deep Learning for Precision Oncology. *Diagnostics* **2022**, *12*.

8. Prakash, U.M.; Iniyar, S.; Dutta, A.K.; Alsubai, S.; Naga Ramesh, J.V.; Mohanty, S.N.; Dudekula, K.V. Multi-Scale Feature Fusion of Deep Convolutional Neural Networks on Cancerous Tumor Detection and Classification Using Biomedical Images. *Sci Rep* **2025**, *15*, doi:10.1038/s41598-024-84949-1.
9. Habeeba, S.; Mahabubullah, Dr.K. Multi-Cancer Detection Using CNN, Inceptionv3, and Vision Transformer (Vit). *International Journal of Innovative Research in Engineering* **2025**, 39–43, doi:10.59256/ijire.20250604006.
10. Nasir, F.; Rahman, S.; Nasir, N. Breast Cancer Detection Using Convolutional Neural Networks: A Deep Learning-Based Approach. *Cureus* **2025**, doi:10.7759/cureus.83421.
11. Jian, W.; Haq, A.U.; Afzal, N.; Khan, S.; Alsolai, H.; Alanazi, S.M.; Zamani, A.T. Developing an Innovative Lung Cancer Detection Model for Accurate Diagnosis in AI Healthcare Systems. *Sci Rep* **2025**, *15*, doi:10.1038/s41598-025-03960-2.
12. Naqvi, M.; Gilani, S.Q.; Syed, T.; Marques, O.; Kim, H.C. Skin Cancer Detection Using Deep Learning – A Review. *Diagnostics* **2023**, *13*.
13. Ghaderzadeh, M.; Aria, M.; Hosseini, A.; Asadi, F.; Bashash, D.; Abolghasemi, H. A Fast and Efficient CNN Model for B-ALL Diagnosis and Its Subtypes Classification Using Peripheral Blood Smear Images. *International Journal of Intelligent Systems* **2022**, *37*, 5113–5133, doi:10.1002/int.22753.
14. Orlov, N. V.; Chen, W.W.; Eckley, D.M.; Macura, T.J.; Shamir, L.; Jaffe, E.S.; Goldberg, I.G. Automatic Classification of Lymphoma Images with Transform-Based Global Features. *IEEE Transactions on Information Technology in Biomedicine* **2010**, *14*, 1003–1013, doi:10.1109/TITB.2010.2050695.
15. Cheng, J. Brain Tumor Dataset. **2017**, doi:10.6084/m9.figshare.1512427.v8.
16. Borkowski, A.A.; Bui, M.M.; Brannon Thomas, L.; Wilson, C.P.; Deland, L.A.; Mastorides, S.M. *Lung and Colon Cancer Histopathological Image Dataset (LC25000)*;
17. Islam, M.N.; Hasan, M.; Hossain, M.K.; Alam, M.G.R.; Uddin, M.Z.; Soylu, A. Vision Transformer and Explainable Transfer Learning Models for Auto Detection of Kidney Cyst, Stone and Tumor from CT-Radiography. *Sci Rep* **2022**, *12*, doi:10.1038/s41598-022-15634-4.
18. Plissiti, M.E.; Dimitrakopoulos, P.; Sfikas, G.; Nikou, C.; Krikoni, O.; Charchanti, A. SIPAKMED: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP); 2018; pp. 3144–3148.
19. Rahman, T.Y.; Mahanta, L.B.; Das, A.K.; Sarma, J.D. Histopathological Imaging Database for Oral Cancer Analysis. *Data Brief* **2020**, *29*, doi:10.1016/j.dib.2020.105114.
20. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Trans Biomed Eng* **2016**, *63*, 1455–1462, doi:10.1109/TBME.2015.2496264.
21. Kim, D.-K. Xception: Deep Learning’s Leap Beyond Inception 2025.
22. Yenigün, O. The Architecture of VGGNet: Breaking Down VGG16 2025.
23. Kundu, N. Exploring ResNet50: An In-Depth Look at the Model Architecture and Code Implementation 2023.
24. Keras. EarlyStopping Callback 2025.
25. TensorFlow; Keras. ReduceLROnPlateau Callback 2025.
26. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis* **2015**, *115*, 211–252, doi:10.1007/s11263-015-0816-y.
27. Sokolova, M.; Lapalme, G. A Systematic Analysis of Performance Measures for Classification Tasks. *Inf Process Manag* **2009**, *45*, 427–437, doi:10.1016/j.ipm.2009.03.002.
28. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. *Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization*;

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.