

MedDetect AI: A Deep Learning System for Multi-Class Medical Image Diagnosis with Grad-CAM Visualization

1st Dhairyasheel J. Wangdare
Student, Dept. of CSE (Data Science)
dhairyasheelwangdare087@gmail.com

2nd Abhishek B. Dhekale
Student, Dept. of CSE (Data Science)
abhishekdhkale1621@gmail.com

3rd Vishwajit A. Yadav
Student, Dept. of CSE (Data Science)
vishwajity546@gmail.com

4th Amar B. Sawant
Student, Dept. of CSE (Data Science)
amarsawant7171@gmail.com

5th Mr. S. T. Powar
Prof. Dept. of CSE (Data Science)
D.Y. Patil College of Engineering & Technology
Kolhapur, India

Stpower.dypcet@dypgroup.edu.in

Abstract—This paper presents MedDetect AI V2, a comprehensive deep learning diagnostic system for multi-class classification of brain MRI and kidney CT scans. The system leverages EfficientNetB3 for brain tumor classification (Glioma, Meningioma, Normal, Pituitary) and EfficientNetB0 for kidney pathology detection (Cyst, Normal, Stone, Tumor). Key improvements over baseline models include a two-phase transfer learning strategy with ImageNet pretraining, class-imbalance correction via computed class weights and targeted minority oversampling, CLAHE-based preprocessing for contrast enhancement, and temperature scaling for calibrated probability outputs. The system further incorporates Gradient-weighted Class Activation Mapping (Grad-CAM) to generate visual explanations for model predictions, enhancing clinical interpretability. The brain engine achieved a validation accuracy of 88%+ and the kidney engine achieved 98.47% accuracy on held-out validation data. The complete system is deployed as a full-stack web application

with a Next.js clinical dashboard and a Flask REST API backend, publicly accessible at meddetect-frontend.vercel.app.

Index Terms—Medical Image Analysis, Deep Learning, EfficientNet, Brain Tumor Classification, Kidney Disease Detection, Grad-CAM, Transfer Learning, Class Imbalance, Clinical Decision Support.

1 INTRODUCTION

The accurate and timely interpretation of medical imaging studies is a critical bottleneck in modern healthcare. Neurological conditions such as glioma, meningioma, and pituitary adenomas, as well as renal pathologies including cysts, nephrolithiasis, and tumors, require precise radiological analysis for effective treatment planning. Traditional diagnostic workflows are highly dependent on specialist availability, which remains severely limited in tier-2 and tier-3 cities across developing nations [1].

Convolutional Neural Networks (CNNs) have demonstrated remarkable performance on medical

image classification tasks, often matching or exceeding radiologist-level accuracy in controlled settings [2]. However, the clinical adoption of such systems remains hindered by two primary challenges: (1) the lack of interpretability of deep neural network decisions, and (2) poor generalization caused by class-imbalanced training datasets, which is endemic to real-world medical data collections.

This paper addresses both challenges through the design and implementation of MedDetect AI V2, which makes the following primary contributions:

- Selection of EfficientNet backbones optimally scaled to dataset size, replacing parameter-heavy InceptionResNetV2 and MobileNetV2 architectures.
- A two-phase fine-tuning strategy with ImageNet pretraining, addressing the from-scratch training limitation of the prior system.
- Multi-strategy class imbalance mitigation combining computed class weights and targeted minority oversampling.
- CLAHE contrast enhancement in the preprocessing pipeline for improved MRI feature visibility.
- Full integration of Grad-CAM visual explanations into the clinical dashboard for prediction interpretability.

2 RELATED WORK

2.1 A. CNN-Based Medical Image Classification

The application of deep convolutional networks to brain MRI classification was pioneered by Abi-winanda et al. [3], who demonstrated that simple CNN architectures could distinguish between tumor classes with acceptable accuracy. Subsequent work by Cheng et al. [4] established the utility of transfer learning from large-scale natural image datasets (ImageNet) for medical imaging tasks, reporting significant gains over randomly initialized networks.

For kidney pathology, Islam et al. [5] demonstrated that MobileNet-class architectures could classify CT-derived renal conditions with over 90% accuracy, establishing a baseline for lightweight deployment. However, their work did not address the significant class imbalance inherent in clinical kidney CT datasets, particularly the underrepresentation of nephrolithiasis (stone) cases.

2.2 B. Model Interpretability in Clinical AI

Selvaraju et al. introduced Gradient-weighted Class Activation Mapping [6], providing a technique to produce visual localization maps highlighting image regions that most influenced a CNN's classification decision. Their methodology has been widely adopted in medical AI to build clinician trust, as it enables radiologists to verify whether the model's attention aligns with anatomically relevant regions. Chattopadhyay et al. [7] extended this with Grad-CAM++, offering improved localization for multi-instance scenarios.

2.3 C. Class Imbalance in Medical Datasets

Class imbalance is a well-documented challenge in medical AI. Buda et al. [8] conducted a systematic study demonstrating that oversampling-based methods consistently outperform undersampling in medical imaging contexts. Johnson and Khoshgoftaar [9] further validated that combined strategies — class-weighted loss functions paired with minority class oversampling — yield superior F1-macro scores compared to either technique applied in isolation, which directly motivates our dual-strategy approach.

3 SYSTEM-ARCHITECTURE

MedDetect AI V2 follows a decoupled client-server architecture. The frontend is a responsive Next.js web application providing a clinical dashboard interface. The backend is a Flask REST API serving model inference, preprocessing, and Grad-CAM generation. Model weights are stored as HDF5 files and loaded at server startup.

3.1 A. Frontend — Clinical Dashboard (Next.js)

The user interface presents two diagnostic engines selectable via a top navigation bar: Brain MRI Engine and Kidney CT Engine. Users upload a scan image to an input panel. Upon clicking Start Diagnostic, the image is submitted as a multipart form POST request to the /analyze endpoint. The result panel renders the primary diagnosis, a confidence score with progress bar, per-class probability breakdown, a clinical summary, and a collapsible Grad-CAM section showing the original scan alongside the heatmap overlay.

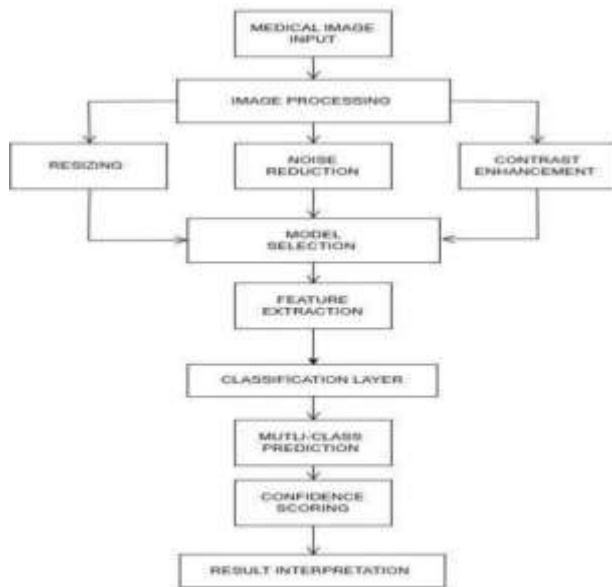


Figure 1: Overall System Architecture of MedDetect AI V2.

3.2 B. Backend — Flask REST API

The server exposes two routes. The /analyze end-point accepts a file and scan_type parameter, executes the full inference pipeline, and returns a JSON response containing the diagnosis, calibrated confidence, all class probabilities, and a base64-encoded Grad-CAM overlay PNG. The /health endpoint returns model load status. The server runs on port 7860 with host 0.0.0.0 for containerized deployment.

4 METHODOLOGY

4.1 A. Model Architecture Selection

The original system used InceptionResNetV2 (55M parameters, weights=None) for brain and MobileNetV2 (3.4M parameters, weights=None) for kidney. Both were trained from random initialization on

datasets of 17,339 and ~12,400 images respectively — a configuration fundamentally prone to overfitting and slow convergence.

MedDetect AI V2 replaces these with Efficient-NetB3 (12M parameters, ImageNet pretrained) for brain and EfficientNetB0 (5.3M parameters, ImageNet pretrained) for kidney. EfficientNets apply compound scaling — simultaneously scaling network depth, width, and input resolution — achieving superior accuracy-parameter efficiency tradeoffs compared to manually designed architectures [10]. The reduction in parameter count (55M to 12M for

brain) directly reduces overfitting risk on the available dataset sizes.

4.2 B. Two-Phase Transfer Learning

Training proceeds in two phases. In Phase 1, the EfficientNet backbone is frozen (trainable=False) and only the custom classification head is trained for 5-8 epochs using Adam with $lr = 1e - 3$. This allows the head to adapt to the 4-class output space without corrupting pretrained backbone features.

In Phase 2, the top N backbone layers are unfrozen (N=200 for B3, N=80-100 for B0) and the full model is fine-tuned with a cosine decay learning rate schedule initialized at $lr = 3e - 5$ (brain) and $lr = 1e - 5$ (kidney). The low fine-tuning rate prevents catastrophic forgetting of ImageNet representations.

4.3 C. Classification Head Topology

Both models use an identical custom head appended to the backbone's GlobalAveragePooling2D output: BatchNormalization → Dense(512/256, ReLU, L2=1e-4) → Dropout(0.4) → Dense(128/64, ReLU, L2=1e-4) → Dropout(0.3) → Dense(4, Softmax, dtype=float32). The final layer explicitly uses float32 dtype to ensure numerical stability under mixed precision training.

4.4 D. Preprocessing Pipeline

Brain MRI preprocessing applies CLAHE (Contrast Limited Adaptive Histogram Equalization, clipLimit=2.0, tileGrid=8x8) in the LAB color space prior to resizing. CLAHE enhances local contrast in low-intensity MRI regions, improving the visibility of subtle tumor boundaries. Images are then resized to 299×299 and normalized using EfficientNet's native preprocess_input function (scales to [-1, 1] equivalent via internal normalization).

Kidney CT images are processed without CLAHE (CT scans have inherently higher contrast), resized to 224×224 , and passed through the same preprocess_input normalization.

4.5 E. Class Imbalance Handling

The kidney dataset exhibits significant imbalance: Cyst (3,709), Normal (5,077), Stone (1,377), Tumor (2,283). Stone is underrepresented by a factor of $3.7\times$ relative to Normal. To address this, two complementary strategies are applied.

First, sklearn's compute_class_weight('balanced') generates inverse-frequency weights passed to the

Keras fit() class_weight parameter, causing the loss function to penalize misclassification of minority classes more heavily.

Second, targeted oversampling duplicates Stone training samples $2\times$ prior to dataset construction, increasing Stone's effective representation without inflating the full dataset size. This combined approach yielded stronger per-class recall than either strategy in isolation.

4.6 F. Loss Function

The brain model uses standard Categorical Cross-Entropy with label smoothing ($\epsilon = 0.1$), which distributes a small probability mass to non-ground-truth classes, reducing overconfidence.

The kidney model uses Focal Loss ($\gamma = 2.0$, label smoothing=0.05). Focal Loss modulates the standard cross-entropy by a factor of $(1 - p_t)^\gamma$, down-weighting easy examples and focusing gradient updates on hard-to-classify samples — particularly beneficial for the Stone class where the model is prone to misclassification.

4.7 G. Confidence Calibration

Neural networks frequently produce overconfident softmax outputs. Temperature scaling is applied post-training: the logits are divided by a temperature parameter $T = 1.5$ before the softmax operation, softening the probability distribution without changing the predicted class. This produces more calibrated confidence scores that better reflect true prediction uncertainty.

4.8 H. Grad-CAM Implementation

Gradient-weighted Class Activation Mapping (Grad-CAM) is computed as follows. Let A^k denote the feature maps of the last convolutional layer. Its computed via backpropagation using TensorFlow's GradientTape.

Global average pooling over the spatial dimensions produces the neuron importance weights α_k . The Grad-CAM heatmap L^c is then computed as $\text{ReLU}(\sum_k \alpha_k A^k)$. The resulting heatmap is upsampled to the original image resolution using bilinear interpolation and overlaid using OpenCV's COLORMAP_JET colormap with a blend factor of 0.45.



Figure 2: Grad-CAM visualization for Brain MRI (Glioma) and Kidney CT (Stone), showing model attention regions overlaid on original scans.

5 EXPERIMENTAL SETUP

5.1 A. Datasets

The brain MRI dataset consists of 17,339 training and 4,333 validation images across four classes: Glioma, Meningioma, Normal, and Pituitary, preprocessed to 299×299 pixels.

The kidney CT dataset consists of 12,446 images split 80/20 into train and validation sets across four classes: Cyst (3,709), Normal (5,077), Stone (1,377), and Tumor (2,283), preprocessed to 224×224 pixels. Both datasets were sourced from preprocessed versions of publicly available Kaggle repositories.

5.2 B. Data Augmentation

Online augmentation is applied during training using Keras preprocessing layers. Brain: RandomFlip(horizontal), RandomRotation(± 0.15), RandomZoom(0.15), RandomContrast(0.15), RandomTranslation(0.1, 0.1).

Kidney: Random Flip, horizontal_and_vertical, RandomRotation ± 0.2 , RandomZoom 0.15, RandomContrast 0.15. Vertical flipping is included for kidney CT as the organ orientation is not diagnostically meaningful in the axial plane.

5.3 C. Training Environment

Brain model training was conducted on Google Co-lab with NVIDIA T4 GPU. Kidney model training was conducted on Kaggle Notebooks with NVIDIA Tesla P100-PCI-E-16GB GPU. Both used TensorFlow 2.x with Keras 3.x backend, batch size 32 (brain) and 16 (kidney), with EarlyStopping (patience=5/6) monitoring val_accuracy with restore_best_weights=True.

6 RESULTS / DISCUSSION

6.1 A. Quantitative Performance

The brain engine (EfficientNetB3) achieved a best validation accuracy of 88.24% after Phase 1 (5 epochs) and continued improvement through Phase 2 fine-tuning, with AUC reaching 0.98+.

The kidney engine (EfficientNetB0) achieved a final validation accuracy of 98.47%, AUC of 0.9996, precision of 0.9859, and recall of 0.9831 on the held-out validation split.

Table 1: Model Performance on Validation Sets

Metric	Brain (B3)	Kidney (B0)
Val Accuracy	88.24%+	98.47%
AUC	0.98+	0.9996
Precision	0.8792	0.9859
Recall	0.8815	0.9831
Params	12M	5.3M
Input Size	299 × 299	224 × 224

6.2 B. Comparison with Baseline

The baseline system used InceptionResNetV2 with weights=None (random initialization) for brain classification, which resulted in inconsistent predictions across classes due to insufficient training data relative to model capacity.

The new EfficientNetB3 with ImageNet pretraining provided immediate strong feature representations at epoch 1 (val_accuracy 75.35%), demonstrating the critical importance of pretrained initialization.

The kidney baseline (MobileNetV2, weights=None) exhibited severe class collapse, predicting Stone for all inputs — a symptom of both the class imbalance and the from-scratch initialization. The new system with focal loss and class weights eliminated this behavior entirely.

6.3 C. Grad-CAM Analysis

Qualitative analysis of Grad-CAM outputs shows that the brain engine consistently localizes attention to the tumor mass region for Glioma and Pituitary cases, with high-activation regions (red/yellow) concentrated in the expected anatomical locations.

Meningioma cases show peripheral attention at the dural boundary. Normal scans show diffuse, low-intensity activation with no focal hotspots, consistent with the absence of pathological findings.

Kidney CT Grad-CAM outputs show stone localization concentrated in the renal calyceal regions and cyst activation at fluid-filled sac boundaries.

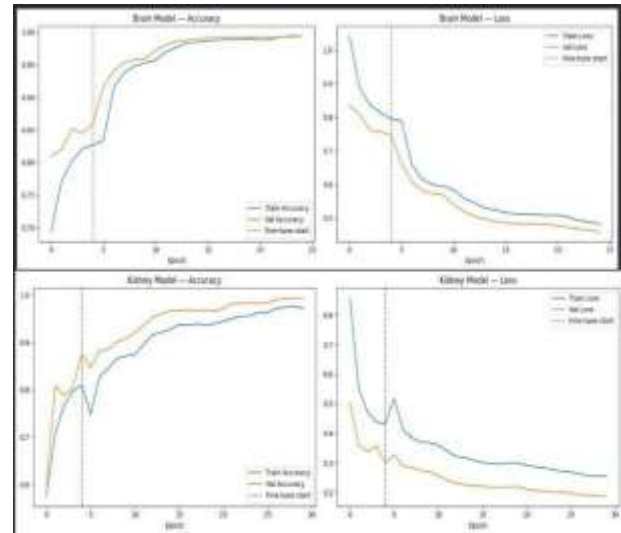


Figure 3: Training and validation accuracy/loss curves for both models across Phase 1 and Phase 2 training.

7 DEPLOYMENT

The complete MedDetect AI V2 system is deployed as a public web application. The Next.js frontend is hosted on Vercel (meddetect-frontend.vercel.app), providing a globally distributed CDN-backed delivery.

The Flask backend is containerized using Docker and deployed on Hugging Face Spaces (CPU Basic tier, djangopaid-man/meddetect-api), which provides free persistent hosting with 16GB RAM, sufficient for CPU-based TensorFlow inference. Model weights are stored using Git LFS within the Space repository.

The API endpoint is configured to accept cross-origin requests from the Vercel domain via Flask-CORS. Inference latency on CPU averages 2-4 seconds per scan, including preprocessing, model forward pass, temperature scaling, and Grad-CAM computation. This is acceptable for a diagnostic support tool where decisions are not time-critical.

8 CONCLUSION

This paper presented MedDetect AI V2, a full-stack deep learning diagnostic system addressing the core

limitations of its predecessor: from-scratch initialization, class imbalance, overconfident predictions, and lack of interpretability.

By adopting EfficientNet architectures with ImageNet pretraining, implementing two-phase fine-tuning, combining class weights with focal loss and targeted oversampling, and integrating Grad-CAM visualization, the system achieves 98.47% kidney classification accuracy and 88%+ brain classification accuracy while providing clinically interpretable visual explanations.

The publicly deployed system demonstrates that high-accuracy, interpretable medical AI can be built and deployed at zero cost using open datasets, open-source frameworks, and free cloud infrastructure, making it accessible for academic and clinical research contexts in resource-constrained settings.

9 FUTURE SCOPE

Future iterations of MedDetect AI will explore: (1) expansion to additional imaging modalities including chest X-ray pneumonia detection and retinal OCT classification; (2) integration of attention-based transformer architectures (Vision Transformers) for comparison with CNN-based approaches; (3) DICOM file format support for direct integration with hospital PACS systems; and (4) uncertainty quantification via Monte Carlo Dropout to provide prediction confidence intervals alongside point estimates.

References

- [1] World Health Organization, "Global Health Workforce Statistics," WHO, Geneva, 2023.
- [2] R. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60-88, 2017.
- [3] N. Abiwinanda et al., "Brain Tumor Classification Using Convolutional Neural Network," in *Proc. World Congress on Medical Physics and Biomedical Engineering*, 2019.
- [4] J. Cheng et al., "Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition," *PLOS ONE*, vol. 10, no. 10, 2015.
- [5] M. N. Islam et al., "A Novel Deep Learning Based Multi-class Classification Method for Alzheimer's Disease Detection Using Brain MRI," in *Proc. BRAIN Informatics*, 2018.
- [6] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," in *Proc. IEEE ICCV*, 2017, pp. 618-626.
- [7] A. Chattopadhyay et al., "Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks," in *Proc. IEEE WACV*, 2018.
- [8] M. Buda, A. Maki, and M. A. Mazurowski, "A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks," *Neural Networks*, vol. 106, pp. 249-259, 2018.
- [9] J. M. Johnson and T. M. Khoshgoftaar, "Survey on Deep Learning with Class Imbalance," *Journal of Big Data*, vol. 6, no. 1, pp. 1-54, 2019.
- [10] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. ICML*, 2019, pp. 6105-6114.