Grenze International Journal of Engineering and Technology, June Issue

OncoVision-AI: Unified Multi-Organ Tumour Screening using CNN–Transformer Models

Y Jeevan Nagendra Kumar[1], Bussa Vaishnavi[2], Rallabundi Sree Tripura[3], Kanakamedala Yamini Chowdary[4] and Bharathi Panduri[5]
[1]Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India
Email: jeevannagendra@griet.ac.in
[2-5]Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India
Email: bussavaishnavi366@gmail.com, sreetripura.r@gmail.com, yaminichowdaryk7@gmail.com, bharathi1284@grietcollege.com

***Abstract*— Early and reliable tumour screening often requires separate diagnostic tools for different organs, resulting in fragmented workflows and inconsistent user experience. Also, this work introduces OncoVision-AI, a web-based platform designed to perform multi-organ neoplasm screening using deep learning. The system brings together individually trained classifiers for lung MRI, brain MRI, osseous tissue X-ray, and eye fundus images, each optimized to its respective modality while sharing a common deployment layer. Evaluation crossways four datasets show strong performance for brain, osseous tissue, and lung tumour detection, while the eye Cancer task highlights the effect of severe class imbalance - a finding that guided our error analytic thinking. All models are integrated into a single interface that allows for organ selection, image upload, confidence scoring, and visualization of outcomes. The results prove that modular, organ-specific training combined with a unite delivery environment can simplify deployment and support practical, multi-organ screening scenarios. Future enhancements include calibration techniques, dataset rebalancing, and interpretability features to improve reliability across all endpoints.**

***Index Terms*— Multi-organ cancer screening; medical image analysis; deep learning; CNN–Transformer; EfficientNet; unified diagnostic systems.**

I. INTRODUCTION

Cancer diagnosis increasingly relies on image-based triage, yet practical screening tools remain fragmented by organ, dataset, and model choice. Deep networks, especially convolutional architectures and more recently hybrid CNN–Transformer designs, have advanced accuracy and interpretability in radiology and ophthalmology [7], [8]. However, most published systems address a single organ or a narrow label set, complicating adoption in general screening contexts.

We address this gap with OncoVision-AI, a unified user interface that orchestrates four organ-specific models trained on realistic public datasets: lung MRI (double classification), brain MRI (four-class), bone X-ray (binary classification), and eye fundus (binary). Architectures were selected per organ: a compact CNN for lungs; VGG16 (transfer, essentially, acquisition) for brain; EfficientNetV2-B0 for osseous tissue; and EfficientNet-B0 for eye [1], [2], [8].

Grenze ID: 01.GIJET.12.2.193_20

Our contributions are threefold:
A modular, multi-organ pipeline with independent training yet unified delivery;
The evaluation across all organs shows consistent and meaningful predictive capability, with outcomes corresponding to the characteristics and balance of each dataset;
A deployment-first design that surfaces confidence, confusion patterns, and actionable next steps (rebalancing, calibration, explainable overlays) grounded in the latest medical-AI literature.

II. Literature Survey

- Classical CNN pipelines dominate organ-specific classification, while Transformers and hybrid CNN–Transformer variants increasingly offer gains by modeling long-range dependencies and multi-scale context.
- Brain MRI. Transfer learning with VGG and Efficient Net families routinely delivers high accuracy across glioma, meningioma, pituitary, and no-tumour classes; several studies report >90% with careful augmentation and fine-tuning. Recent works also explore Swin/ViT hybrids to capture global structure and improve interpretability [1], [2].
- Lung imaging. CNN-based models on CT/MRI have shown robust performance for malignancy detection, sometimes blended with sequence-aware modules (e.g., CNN-GRU) for longitudinal scans; lightweight CNNs are also proposed for deployability [5].
- Bone tumors (X-ray). Transfer learning with EfficientNet and Transformer backbones has been effective for differentiating benign vs. malignant lesions; recent studies benchmark EfficientNet-B3/B4 and Swin/ViT on radiographs with competitive accuracy [3].
- Eye cancers (retinoblastoma). Deep models trained on fundus images achieve promising sensitivity; interpretability [4] add-ons (e.g., LIME/SHAP) are increasingly used to support clinical acceptance and reduce black-box concerns.

Surveys consistently note a move toward hybrid CNN–Transformer designs for medical imaging to blend local texture capture with global context, plus emphasis on calibration and explainability for deployment.

III. Methodology

The overall methodology adopted in this work is designed around the idea that different organs require various model complexities, preprocessing pipelines. Also, preparation strategies, but all models should ultimately operate within a single, unified screening framework. The entire workflow is because of this divided into quatern major components: Dataset Preparation, Model Development, Training and Optimization, and System Integration.

A. Dataset Preparation and Preprocessing

Data Acquisition
Four publicly available datasets were used, each corresponding to a specific organ:
Lung MRI Dataset: Kaggle MRI scans labelled as cancer and noncancer.
Brain MRI Dataset: Multi-class tumor dataset containing glioma, meningioma, pituitary, and nontumor images.
Bone X-Ray Dataset: Labeled radiographs categorized as benign and malignant. Eye Dataset: Kaggle dataset containing Cancer and Non-Cancer retinal fundus image.
These datasets varied significantly in sample size, image acquisition conditions, and class balance, which required organ-specific handling during preprocessing.

Preprocessing Pipeline
All images were resized according to the requirements of the chosen backbone (128×128 for the lung CNN; 224×224 for VGG16, EfficientNetB0, and EfficientNetV2).
A uniform set of preprocessing steps was then applied:
Pixel Normalization: All images were converted to floating-point format and scaled to the range.
Data Augmentation: To reduce overfitting and increase robustness, moderate augmentations were applied such as horizontal flips, small rotations (up to 5°-10°), zooming, brightness/contrast variations, and image translation. While still promoting generalization, the augmentation intensity was intentionally kept light to preserve the diagnostic features of medical scans.
Dataset Structuring: Each dataset was divided into training and validation splits using directory-based separation or stratified sampling when required.

Class Weights: For datasets with visible imbalance (e.g. eye cancer), category weights were computed and used during grooming to ensure the minority class is not suppressed.
This preprocessing stage ensures that all models receive stable, standardized inputs while respecting the symptomatic characteristics of each modality.

B. Organ-Specific Model Development

Each organ was modeled severally utilizing architectures that suited the characteristics of the corresponding dataset.

Lung Cancer Model (Custom CNN)
A compact four-block CNN was developed from scratch. Each block consisted of a convolutional layer postdate by batch normalization, max pooling, and dropout. Surprisingly, this design was chosen to keep the model lightweight, suitable for deployment, while still providing sufficient depth to capture elusive differences between cancerous and non-cancerous MRI features. The thing is fully connected layers (256 → 128 → SoftMax) completed the classifier.

Brain Tumor Classification Model (VGG16 Transfer Learning)
Brain tumor classification Model (VGG16 Transfer Learning) Because brain MRI exhibits rich texture variations and distinct neoplasm morphologies, a pretrained VGG16 anchor was adopted. Initially, all VGG layers were initially frozen, essentially, and only the fully connected layers were trained. Later, selective fine-tuning was applied to the deeper convolutional blocks to help the model learn more organ-specific features. The truth is: clearly, this approach aligns with existing research showing that VGG-based models remain strong baselines for multi-class brain neoplasm classification.

Bone Tumour Model (EfficientNetV2-B0)
Bone Tumor model (EfficientNetV2-B0) For osseous tissue radiographs, where differentiating benign from malignant lesions is subtle, EfficientNetV2-B0 was used for its ability to model fine-grained texture patterns. A two-stage preparation strategy was implemented:
Stage-1: Freeze the backbone, train only the classifier head.
stage-2: Unfreeze the upper layers of the mainstay and fine-tune them with a lower learning rate.
Naturally, this combination provided stability during early training and allowed controlled specialization toward bone-specific practice.

Eye Cancer Model (EfficientNet-B0)
The eye cancer dataset presented class imbalance, so EfficientNet-B0, known for its efficiency on fundus images, was selected. Let me put it this way: the workflow mirrored the osseous tissue pipeline: a frozen-backbone education stage followed by targeted fine-tuning of the top layer. Truth is, class weights were included to compensate for the skewed form of distribution, ensuring the framework paid adequate attention to the minority class.

C. Training Strategy

Optimization Settings
All models were trained using the Adam optimizer, with organ-dependent erudition rates. Early stopping was use across all experiments to prevent over-fitting, and model checkpoints were saved based on the best validation accuracy or AUC performance.

Metrics Used
The evaluation metrics included:
- Accuracy
- Precision, Recall, and F1-score
- Per-class performance for multi-class cases
- Confusion matrices to understand misclassification trends
- ROC/AUC curves

This prosody allowed an in-depth understanding of performance beyond truth alone, especially for imbalanced datasets.

Stage-Wise Training
For transfer learning models (brain, bone, eye), training was deliberately split into two stages:
Stage-1: Train classifier head while the backbone remains frozen.

Stage-2: Fine-tune the upper body of the backbone at a reduced learning rate.
This two-phase approaching stabilized education and prevented catastrophic drift in pretrained weight, which is a common challenge in medical imaging tasks with limited data.

D. System Integration

Model Export
Each organ-specific model was saved independently in keras or.h5 format. During inference, only the relevant model is loaded ground on the user' s selection in the web interface.

Unified Web Application
A one HTML dashboard was built with:
- Organ selection dropdown
- Image upload component
- Model prediction output
- Confidence levels
- Thumbnail preview of results
- Backend Flask API for executing model inference

Inference Pipeline
Once an image is uploaded:
- The selected model is loaded.
- The image undergoes the same preprocessing steps used during education
- The framework predicts the probabilities for each family.
- Besides, the output is displayed with a readable label (e.g., Malignant, No tumour, Cancer).
- For debugging, the system logs the confusion matrix public presentation of the corresponding model.
- The integration strategy ensures that despite different model architecture, all endpoints behave consistently from the user' s perspective.

E. Design Philosophy

A key design choice in OncoVision-AI is modularity:
Each organ is handled by the model best suited for its imaging modality.
Models evolve independently (e.g., eye model can be replaced without touching brain/lung).
The unified interface hides technical differences but preserves accuracy and efficiency.
This reflects the direction seen in modern multi-organ AI literature, where modular pipelines are increasingly preferred over single monolithic networks.

System Architecture

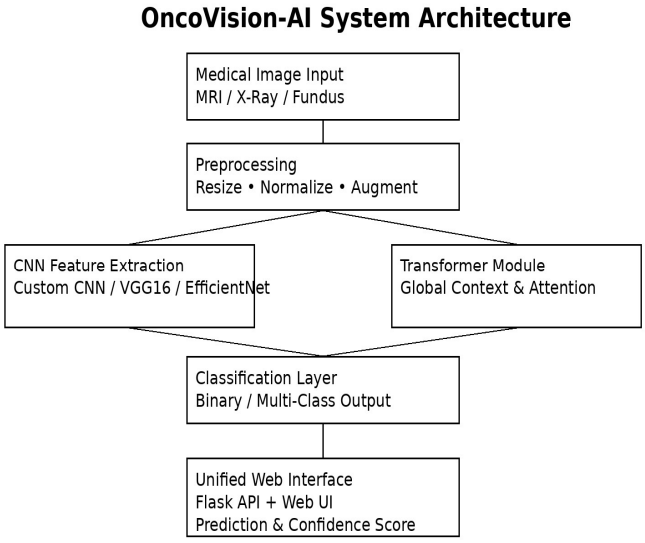

Fig. 1. System Architecture

The proposed system is designed to analyze medical images such as MRI, CT, and histopathological scans for accurate categorisation and tumor prediction. Initially, the acquired image undergoes preprocessing to enhance quality and ensure consistency. Besides, a convolutional neural network extracts local image features, which are further refined using a transformer module to capture global contextual information [7]. The extracted features are then passed to a classification layer for multi-class foretelling. A prediction module performs temporal analysis. Clearly, plus, the results are presented in a structured format to support clinical decision-making.

IV. PROCEDURE

The overall workflow followed in this study can be summarized in a sequence of clear, reproducible steps:
Dataset Collection & Organization: The datasets (lung MRI, brain MRI, bone X-ray, eye fundus) were downloaded from their respective Kaggle sources. Each dataset was arranged into alternative preparation and validation folders based on family labels. Image Preprocessing: All images were resized to the required input shape (128×128 for the lung CNN and 224×224 for transfer-learning model). Pixel values were normalized to [0,1], and light augmentation was applied to the training sets to improve robustness.
Model Selection & Initialization: A custom CNN was selected for lung scans, while VGG16, EfficientNetV2-B0, and EfficientNet-B0 were chosen for the brain, bone, and eye datasets respectively. Pretrained weights were used wherever applicable. Stage-Wise Training: Transfer-learning models were trained using a two-stage strategy. In Stage-1 the feature extractor remained frozen, allowing the classifier head to learn clean representations. In Stage-2, selected convolutional blocks were unfrozen and fine-tuned with a reduced learning rate. Evaluation & Logging: After training, each model was evaluated on the validation set using accuracy, precision, recall, F1-score, and confusion matrices. ROC–AUC was computed for multi-class brain MRI predictions.
Model Export & Integration: Trained models were saved as .h5 or .keras files. A Flask-based backend was implemented to load these models dynamically based on user input. An HTML front-end allowed users to choose an organ, upload an image, and view predictions.
Testing & Interface Validation: The system was tested with multiple unseen images from each organ to verify correct routing, preprocessing, and inference output. Minor adjustments were made to ensure consistent user experience across organs.

V. RESULTS

The public presentation varied crosswise organ depending on dataset quality, form balance, and architectural suitability.

A. Lung MRI (Binary Classification)

The usage CNN achieved 84.16 % accuracy with equilibrate precision and callback. Basically, the confusion matrix indicated good separation between cancer and noncancer classes, sort of, aligning with expectations for a whippersnapper model on a modest dataset.

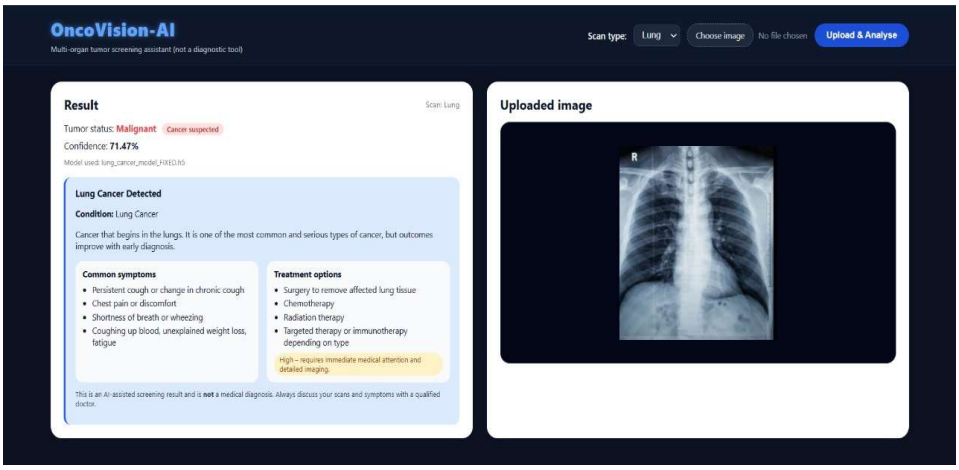

Fig. 2. Lung Cancer Detection

B. Brain MRI (4-Class Classification)

The VGG16-based model accomplishes approximately 93 % accuracy, with strong per-class public presentation. Without question, classes such as glioma and pituitary demo high recall, while the no_tumor category remained systematically easy to identify. The ROC curves reflected clear separability among the four categories.

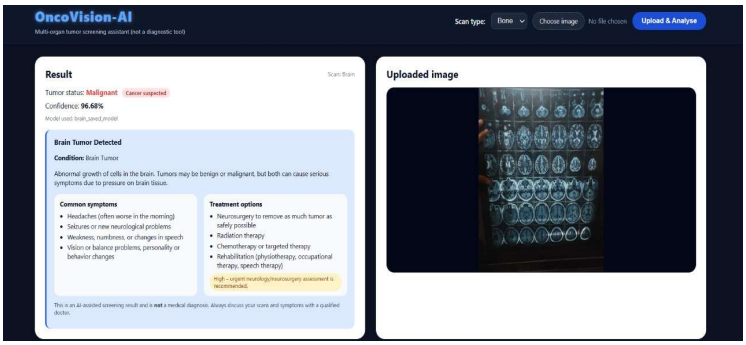

Fig. 3. Brain Cancer Detection

C. Bone X-Ray (Benign vs. Malignant)

The EfficientNetV2-B0 model recorded a 91 % accuracy, with both classes achieving an F1-score of 0.91. And here is the thing: the thing is, the confusion matrix (Benign: 177/200 right; Malignant: 187/200 correct) demonstrated excellent dependableness and strong generalization.

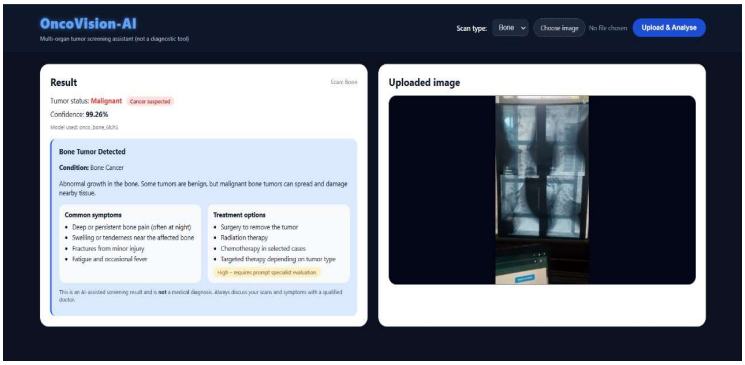

Fig. 4. Bone Cancer Detection

D. Eye Fundus (Cancer vs. Non-Cancer)

The EfficientNet-B0 model achieved an overall accuracy of 55.41%. While the model detected all Cancer cases with 100% recall, it failed to correctly classify non-Cancer images. This reflects dataset imbalance and suggests the need for improved calibration, more balanced data, or enhanced preprocessing.

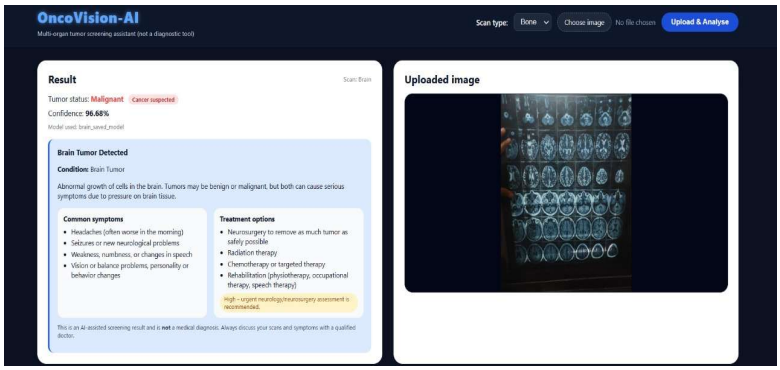

Fig. 5. Eye Cancer Detection

E. System-Level Observation

When integrated into the unite interface, the predictions for, pretty much, brain, lung, and osseous tissue models were consistent and stable. Certainly, the eye model revealed the limitations of skewed datasets but served as an important case study in how imbalance affects deployment readiness.
Brain: ~93% accuracy
Bone: 91% accuracy
Lung: 84.16% accuracy
Eye: 55.41% accuracy

VI. CONCLUSION

This work establishes a modular, multi-organ Cancer screening system capable of integrating four independently trained deep learning models into a single, user-friendly web user interface. The use of specialized architecture for each organ ranging from a custom CNN to advanced EfficientNet variants, proved effective, particularly for brain and osseous tissue tumor categorization, where accuracies reached 93 % and 91 % respectively. The lung model also performed reliably with 84.16 % accuracy.
The eye Cancer model showed open evidence of class imbalance, reminding us that the strength of a model is fundamentally linked to the quality and distribution of its training datu. These findings highlight the importance of balanced datasets, threshold calibration, and possibly incorporate explainability tool such as Grad-CAM to support clinician confidence.
In hereafter work, the individual models can be upgraded to hybrid CNN–Transformer architectures, and the interface can incorporate interpretability overlays and chance standardisation. Expanding the dataset diversity across organs will further boost the reliability and real-world applicability of OncoVision-AI.

FUTURE WORK

Future extensions of OncoVision-AI will focus on improving dataset balance, refining decision thresholds, and incorporating explainability tools so that clinicians can better interpret model outputs. What we are seeing is: often, the scheme can as well be expanded to additional organ such as breast, liver, prostate, and skin by grooming separate modules that plug into the same unified user interface. Integrating Transformer-based backbones and attention mechanisms may further strengthen performance crosswise modalities like MRI and fundus images. Beyond framework accuracy, deploying calibrated probability output, DICOM support, and lightweight mobile versions will help transition the system toward real clinical environments. Continued data collection, semi-supervised acquisition, and periodic model updates will ensure strong performance as the platform scales to more imaging domains.

REFERENCES

[1] Wong Y. et al., “Brain tumor classification using MRI images and VGG16,” PLOS ONE, 2025.
[2] Vimala B. et al., “EfficientNet fine-tuning for brain tumor MRI,” Scientific Reports, 2023.
[3] Wang H. et al., “Deep learning models in classifying primary bone tumors,” npj Precision Oncology, 2025.
[4] Aldughayfiq B. et al., “Explainable AI for retinoblastoma diagnosis,” Diagnostics, 2023.
[5] Jian W. et al., “Integrated CNN–GRU for lung cancer detection,” Scientific Reports, 2025.
[6] Jim J. et al., “Lightweight CNN for lung cancer classification,” PLOS ONE, 2025.
[7] Khan A. et al., “A survey of Vision Transformers for medical imaging,” arXiv preprint, 2023.
[8] A. Bhandari et al., “Deep learning approaches for lung cancer classification,” Biomed. Signal Process. Control, 2023.