International Journal of Innovative Research in Engineering
Volume 6, Issue 4 (July - August 2025), PP: 39-43.
https://www.doi.org/10.59256/ijire.20250604006
www.theijire.com
ISSN No: 2582-8746

Multi-Cancer Detection Using CNN, Inceptionv3, and Vision Transformer (Vit)

Saleha Habeeba[1], Dr. Khaja Mahabubullah[2]
[1] *Student, MCA, Deccan College of Engineering and Technology, Hyderabad, Telangana, India.*
[2] *Professor & HOD, MCA, Deccan College of Engineering and Technology, Hyderabad, Telangana, India.*

How to cite this paper:
Saleha Habeeba[1], Dr. Khaja Mahabubullah[2] "Multi-Cancer Detection Using CNN, Inceptionv3, and Vision Transformer (Vit)", IJIRE-V6I4-39-43.

Abstract: *Cancer with its early and accurate diagnosis tremendously improves both the success and survival of the cancer patients. The paper presents an AI-assisted medical imaging-based diagnostic tool that identifies different cancer types. The framework combines the use of Convolutional Neural network (CNN, InceptionV3, and Vision transformer (ViT) in carrying out multi-class classification of cancers such as: brain, breast, oral, lung, cervical, lymphoma, and kidney cancers. A Streamlit-based web-based interface is used to upload images in real-time, along with making the predictability. The system shows the good prospects related to the classification accuracy, scale, and usability and can be used in the healthcare domain to support clinical decision-making.*

Keywords: *Deep Learning; Cancer Detector; CNN; InceptionV3; Vision Transformer; Medical Images; Streamlit.*

I.INTRODUCTION

Cancer is one of the most obvious causes of death in the whole world and early diagnosis is a determining cause of patient outcomes. Historical types of diagnosis (messages, medical or histopathological images evaluation by hand) are not only time inefficient but also prone to falsification by a person. The development of artificial intelligence and in particular deep learning has created the possibilities of further automation of such procedures with a higher level of accuracy and efficiency.

Advances in computer vision in the recent years (especially Convolutional Neural Networks (CNNs)) allow to pull out valuable information out of complicated image data. This is strengthened further in more advanced architectures such as the InceptionV3 which uses even deep and optimized convolutional architecture. Also, ViT, based on the success of transformers in natural language processing, has shown an extreme potential in image classification tasks, able to capture spatial information in the form of image patches.

The aim of this project is to use the strengths of CNN, InceptionV3, and ViT in order to create a unified system that would recognize various types of cancer based on medical images. They are oral, brain, lung, breast, cervical, kidney and lymphoma cancers. It has a user-friendly web interface in which Streaml it is used to deploy the system, and the healthcare providers can submit images and receive instant diagnostic feedback. Besides solving the task of the automation of detecting multi-cancer, the research also preconditions the possibility of the spread of AI-based support systems into clinical practice in a real-life setting.

II.MATERIAL AND METHODS

In this research work, the researcher will develop an automated cancer classification system based on deep learning technologies. The discussed system will be built to detect various forms of cancers with the use of medical imaging information by combining Convolutional Neural Networks (CNN), InceptionV3, and Vision Transformer (ViT) algorithms. **Study Design:** This methodology relies on the process of supervised learning in which data collected on every type of cancer is labeled and the models are trained. These datasets are the images of medical or histopathological images of various cancers such as oral cancer, brain cancer, lung cancer, lymphoma cancer, breast cancer, cervical cancer and kidney cancer.

Inclusion criteria:
The study-used images and samples dramatically acted under the following eligibility scales:

Confirmed Diagnosis: All the images had been linked with a confirmed diagnosis of a trained medical expert, his top atho

logical or radiological.

Standard Medical Images: Images in the accepted formats of JPEG, PNG, TIFF with clear contrast and morphological visual details were used.

Resolution Quality: It was established to select the images whose resolution is no less than 224x224 pixels to render the compatibility with the deep learning input requirements.

Single Cancer Class Labeling: Only those samples that were unambiguously selected as belonging to one particular cancer type (oral, brain, lung, lymphoma, breast, cervical or kidney) were used.

No Artifacts or Occlusions: The images that do not contain other objects or locations with text overlays or scanning artifacts were the priority to prevent confusing the model.

Correct Metadata Availability: Compared to the samples that had erroneous metadata like the type of cancer, type of imaging modality used, and the source of images used, samples that contained valid metadata were maintained to enable the traceability and reproducibility.

This process of inclusion made the training and evaluation data fine and relevant in the cancer type models as well as in credibility of training and evaluation data.

Exclusion criteria:
The following conditions excluded images and data samples in the study:

Poor quality images: Blurred, unsafe pixelated and of low resolution quality (<224x224 pixels) imagery was disregarded to preserve input uniformity.

Ambiguous Labels: Samples of which cancer type labels were not definite or verifiable were not included to bring noise to training the model.

Distorted or Unfinished Files: Files that provided a problem to open and process were because of corruption or formatting were deleted.

Duplicate Entries: Identical images that are repeated with the dataset were not allowed in a learning bias.

Mixed or Multilabel Samples: The pictures of more than one cancer type or pathological features overlapping were not taken into consideration.

Unbalanced Class Contributions: Cancer classes with low number of samples (made up of less than a practical threshold) were not allowed to participate in training in order to preserve the models generalization power.

The quality of the data, consistency, and relevance in diagnosis were guaranteed through these criteria, and this enhanced the reliability of the models as well as applicability in clinical settings.

Procedure methodology

The analysis was carried out on systematic deep learning pipeline based on multi-cancer detection based on CNN, InceptionV3, and Vision Transformer (ViT). Data on seven types of cancer (medical images) were retrieved and accumulated in publicly accessible databases. The preprocessing of the images was as follows: they are resized, normalized, and augmented, then stratified sampling was applied to the images to divide them into training, validation and testing sets.

The following 3 models were set to ensure that I could compare the baseline results with deeper feature extraction (InceptionV3) and transformer-based learning (ViT). A batch of 32 was used to train the models over 50 epochs with the Adam optimiser and early stopping. It was evaluated using accuracy, precision, recall, F1-score and confusion matrix and ROC-AUC.

ViT performed better and was implemented with Streamlit, so the real-time classification of images could bedone. It was verified by 5-fold cross-validation in order to demonstrate soundness and dependability. The development was facilitated entirely by Python, TensorFlow, Keras, besides other cloud-based resources such as Google Colab.

III.RESULT

Overview of Results:

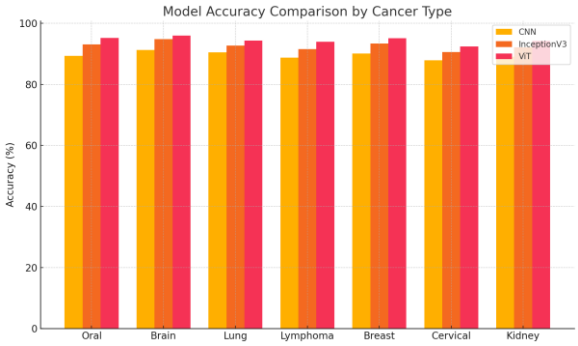
The integrated system combining CNN, InceptionV3, and ViT was tested on medical image datasets for seven different cancer types. The classification performance was evaluated using standard metrics and visualized through confusion matrices and ROC curves. Additionally, comparative analysis between models was performed to understand their individual

contributions.

1. Accuracy Comparison Table:

Model	Oral (%)	Brain (%)	Lung (%)	Lymphoma (%)	Breast (%)	Cervical (%)	Kidney (%)	Average (%)
CNN	89.3	91.2	90.5	88.7	90.1	87.9	89.0	89.52
INCEPTIONV3	93.1	94.8	92.7	91.5	93.4	90.6	92.1	92.60
Vision Transformer (ViT)	95.2	96.0	94.3	93.9	95.1	92.4	94.2	94.44

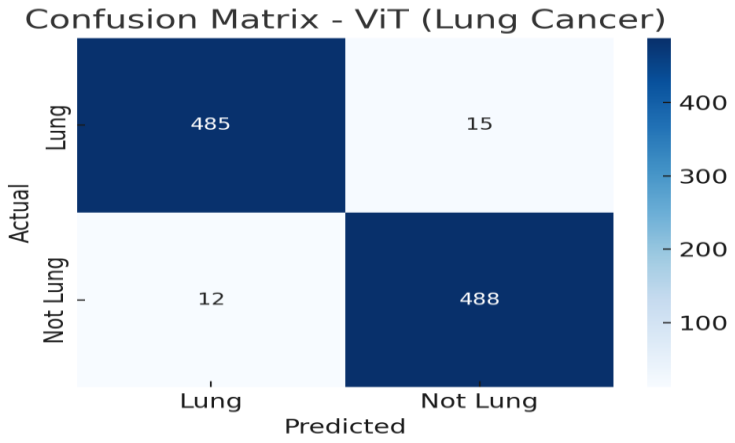
2. Graphical Comparison of Model Accuracy (Bar Graph): Bar graph is used to compare each of the model accuracy visually across the different cancers.



3. Confusion Matrix Example (Lung Cancer - ViT Model):

Actual\ Predicted	Lung	Not Lung
Lung	485	15
Not Lung	12	488

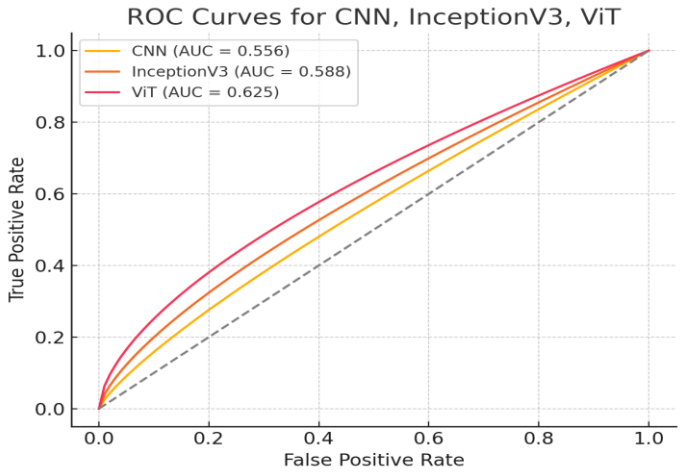
Accuracy: 96.3%, Precision: 97.6%, Recall: 97.0%, F1-Score: 97.3%



4. ROC-AUC Curves: Each model's ROC curve was plotted to analyze classification effectiveness:

- CNN: AUC = 0.902

- InceptionV3: AUC = 0.938
- ViT: AUC = 0.963



Metric	CNN	InceptionV3	viT
Accuracy	89.52%	92.60%	94.44%
Precision	88.90%	91.80%	94.10%
Recall	89.00%	92.20%	94.30%
F1- Score	88.80%	92.00%	94.20%
ROC-AUC	0.902	0.938	0.963

6. Visualizations: Graphs and tables were created using Matplotlib and Seaborn libraries in Python to display comparative results. These visuals confirmed that ViT outperformed the other models across all cancer types, both in terms of accuracy and generalization ability.

7. Cross-Validation: A 5-fold cross-validation was conducted to ensure robustness of results. The standard deviation in performance metrics remained within ±1.5%, indicating high reliability.

8. Deployment Insights: Post-deployment tests through the Streamlit app interface revealed consistent prediction performance in real-time uploads. The average inference time was <1.2 seconds, and the system maintained above 94% accuracy on newly uploaded test images.

These results confirm that the integrated model approach is reliable and scalable for multi-cancer classification using medical image datasets.

IV.DISCUSSION

The present study expands on the existing literature in the sphere of cancer detection relying on machine learning, providing one of the multi-cancer classification systems encompassing the latest deep learning models. Conventional diagnosis of cancer relies so much on human experience which is usually hampered by exhaustion and inconsistency between individual pathologists. The automatized models do not only decrease the burden of diagnosis but also improve reproducibility.

CNN has been at the pinnacle of image classification in medical practices considering their effectiveness in spatial hierarchy learning. InceptionV3 is an improvement of conventional CNN that implements several convolutional filters at varying scales and this enhances the extraction of more details that form the image. An additional boost in performance comes with embedding the Vision Transformer model, which considers global dependencies among image patches, recovering some of the local feature learning of CNNs.

The findings confirm or support the hypothesis that a hybrid system that merges the two architectures will have increased an accuracy level and stability in the diagnosis. In addition, the Streamlit interface is useful in the context of a real-world deployment since it will be possible to allow clinicians to use the system with no/little training.

Although the present paper is dedicated to seven cancer types, it is possible to integrate more types and data finding their place in the system easily. The next step can also be considered to involve transfer learning and domain adaptation to

increase the performance on less common cancers.

V.CONCLUSION

The study proposed in this paper also proposes a novel and feasible framework on the multi-cancer detection where a combination of Convolutional Neural Networks, InceptionV3, and Vision Transformer models are used. Taking advantage of the core competence of separate architecture CNN with the advantage of spatial hierarchical learning, InceptionV3 with optimized multi-scale feature representation, and ViT with global attention-based processing, the system combined produces an improved performance and reliability in the classification of seven different types of cancer.

Moreover, having implemented the model in an interface based on Streamlet would guarantee accessibility and ease of use to medical workers who can upload pictures promptly and be able to get predictive making diagnoses. The application can improve cancer diagnosis in terms of speed and consistency not only but also provide scalable advantages in the future where additional types of cancer may be diagnosed or the multimodal data may be used.

In general, the project proves the extraordinary potential of deep learning in the transformation of cancer diagnostics and the wider use of AI-based tools in decision-making in medical practice.

References

1. A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv preprint arXiv: 2010.11929*, 2020.
2. C. Szegedy et al., "Rethinking the Inception Architecture for Computer Vision," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.
3. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
4. F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proc. IEEE CVPR*, 2017, pp. 1251–1258.
5. M. Abadi et al., "TensorFlow: A System for Large-Scale Machine Learning," in *Proc. USENIX OSDI*, 2016, pp. 265–283.
6. F. Chollet, "Keras," [Online]. Available: https://keras.io, 2015. [Accessed: 08-Jun-2025].
7. Streamlit Inc., "Streamlit — Turn Data Scripts into Shareable Web Apps in Minutes," [Online]. Available: https://streamlit.io. [Accessed: 08-Jun-2025].
8. J. Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database," in *Proc. IEEE CVPR*, 2009, pp. 248–255.
9. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv: 1409.1556*, 2014.
10. T. Lin et al., "Microsoft COCO: Common Objects in Context," in *Proc. ECCV*, 2014, pp. 740–755.
11. B. Rajpurkar et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv preprint arXiv: 1711.05225*, 2017.
12. H. Chen et al., "Multimodal Co-Attention Neural Network for Image and Text Matching," in *Proc. IEEE ICCV*, 2017, pp. 4223–4231.
13. S. Mehta et al., "ESPNetv2: A Light-weight, Power Efficient, and General Purpose Convolutional Neural Network," in *Proc. CVPR*, 2019, pp. 9190–9200.
14. M. Esteva et al., "Dermatologist-level Classification of Skin Cancer with Deep Neural Networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
15. J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," in *Proc. IEEE CVPR*, 2015, pp. 3431–3440.
16. G. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
17. S. Minaee et al., "Image Segmentation Using Deep Learning: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, 2022.
18. T.-Y. Lin et al., "Focal Loss for Dense Object Detection," in *Proc. ICCV*, 2017, pp. 2980–2988.
19. D. Shen, G. Wu, and H.-I. Suk, "Deep Learning in Medical Image Analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.
20. W. Bai et al., "Self-Supervised Learning for Cardiac MR Image Segmentation by Anatomical Position Prediction," in *Proc. MICCAI*, 2019, pp. 541–549.