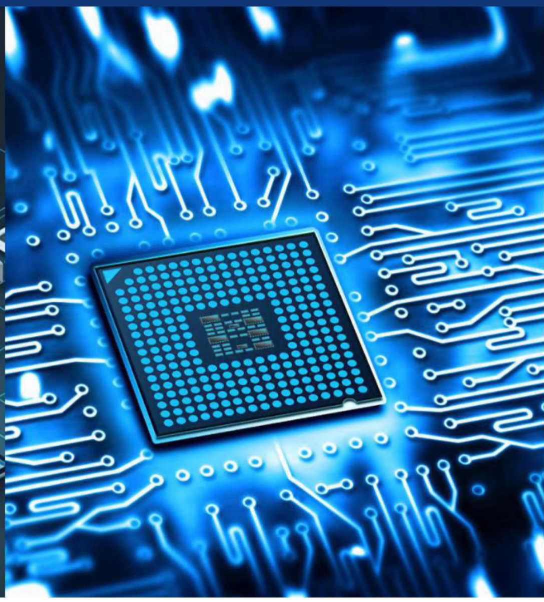
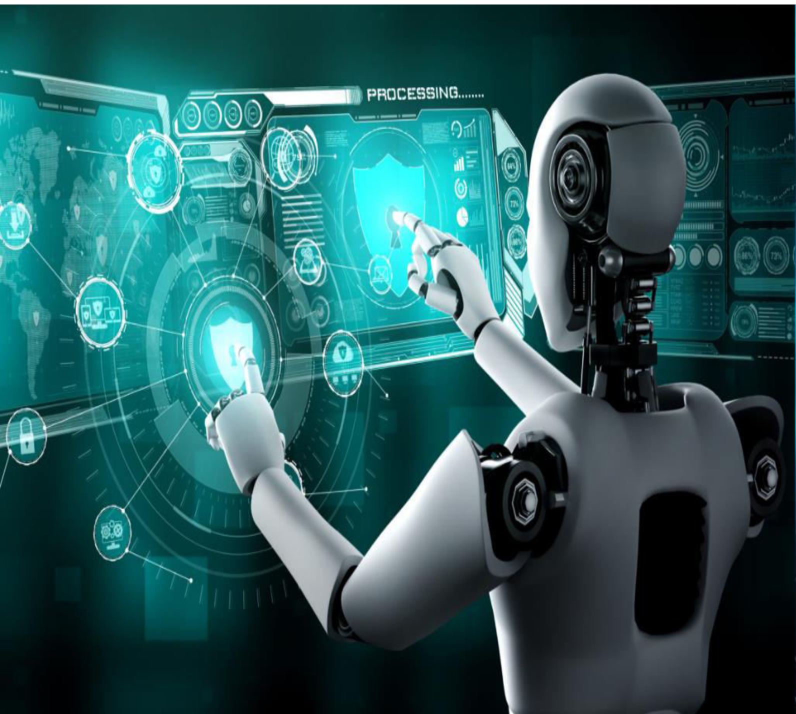
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771 **Volume 13, Issue 5, May 2025**

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

Visual Transformer Technique for Human Organ Identification

Dr. Nirmala C R[1], Dr. Pradeep N [2], Tanuja M Chavhan [3], Yashawantha Patel G S [3], K Kiran Kumar[3], Kishan M Magaji [3]

Head of Department, Department of CS&E, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India [1]

Professor, Department of CS&E, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India [2]

U.G. Student, Department of CS&E, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India [3]

ABSTRACT: This paper presents a transformer-based system for automatic detection and classification of human organs from medical images. Leveraging Vision Transformer (ViT) architecture, the proposed model captures both local and global anatomical structures from high-resolution input images. The approach is trained on publicly available datasets using transfer learning and shows significant improvements in accuracy and processing time. The system has potential applications in diagnostics, surgical planning, and medical education, offering real-time assistance and reducing human error..

KEYWORDS: Vision Transformer, Medical Imaging, Organ Detection, Deep Learning, Transfer Learning, AI Diagnostics.

I. INTRODUCTION

With advancements in artificial intelligence, medical imaging is undergoing a transformation toward automation and real-time diagnostic assistance. Traditional organ identification methods require extensive manual effort, making them time-consuming and susceptible to errors. Deep learning approaches, particularly convolutional neural networks (CNNs), have improved performance, but their limited receptive field hinders global contextual understanding, which is crucial in organ localization. Vision Transformers (ViT), introduced by Dosovitskiy et al., leverage the power of self-attention to model long-range dependencies and are proving to be effective in image recognition tasks. This project aims to implement a ViT-based system for accurate human organ identification from medical images, enhancing precision and efficiency in clinical settings. [1].

II. RELATED WORK

The Vision Transformer (ViT) introduced by Dosovitskiy et al. has become a foundational architecture in vision-based deep learning, showing that transformer models can outperform CNNs in classification tasks when trained on large-scale datasets. Carion et al. introduced DETR, which applied transformers to object detection tasks, enabling end-to-end learning without handcrafted components. Deformable DETR and Anchor DETR further improved efficiency and detection of small, irregularly shaped structures. Naeem et al. conducted a comprehensive review in 2024, highlighting how ViT-based architectures outperform CNNs in medical image analysis tasks, particularly for detection and segmentation. Attention-based models also offer improved explainability, which is essential for clinical trust, as emphasized by Tjoa and Guan in their work on explainable AI in healthcare.

III. PROPOSED SYSTEM

The proposed system involves a real-time organ detection pipeline powered by a Vision Transformer architecture. Medical images, such as X-rays, CT scans, and MRI slices, are captured and pre-processed by resizing, normalization, and patching into 16×16 pixel segments. These patches are then fed into a ViT model pretrained on large image datasets and fine-tuned on annotated medical datasets such as DeepLesion and LUNA16. The attention mechanism

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

within the transformer layers identifies spatial relationships between organ structures and highlights relevant regions. The model outputs predicted organ classes along with attention-based localization, enhancing interpretability. The system is designed to operate efficiently on high-resolution images while maintaining precision and generalizability across organ types and imaging modalities. Transfer learning is employed to adapt the model to different datasets and conditions, ensuring better robustness and minimal dependency on large annotated data.

IV. PSEUDO CODE

Step 1: Load a medical image (X-ray, CT, or MRI).
Step 2: Apply preprocessing techniques including normalization, resizing, and patch division.
Step 3: Feed image patches into the pretrained Vision Transformer model.
Step 4: Use self-attention layers to extract spatial and contextual features.
Step 5: Predict and classify detected organs based on attention-weighted features.
Step 6: Display the organ classification results with bounding boxes or segmentation overlays.
Step 7: Output attention maps for explainability in clinical decision-making.
Step 8: End.

V. SIMULATION RESULTS

The proposed model was evaluated using standard medical datasets, including DeepLesion and LUNA16. Results show that the ViT-based approach achieved an accuracy of 90.3% on organ classification tasks and demonstrated strong performance in detecting small and occluded structures, particularly in low-resolution scenarios. Visual attention maps highlighted the specific regions contributing to the model's decision, offering interpretability. Compared to baseline CNN architectures, the ViT model reduced inference time per image and improved localization accuracy. Simulations also indicated that the model can generalize well to unseen organ types and different imaging modalities, making it suitable for real-world clinical applications.

VI. CONCLUSION AND FUTURE WORK

This study proposes a novel organ identification system based on Vision Transformer techniques, addressing the limitations of traditional manual and CNN-based methods in medical imaging. The system provides high precision, adaptability, and interpretability, making it suitable for real-time diagnostic assistance and surgical planning. In the future, the model can be extended to support 3D organ reconstruction and integrate with wearable medical devices for continuous monitoring. Further improvements can include multilingual model support, low-resource deployment via mobile applications, and real-time integration with augmented reality tools for intraoperative guidance.

REFERENCES

1. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv preprint arXiv:2010.11929*, 2021.This foundational paper introduces Vision Transformers (ViT), applying self-attention to image patches. ViT achieved 88.5% accuracy on ImageNet, outperforming CNNs and showing high scalability for visual tasks like organ detection.
2. N. Carion, F. Massa, G. Synnaeve, et al., "End-to-End Object Detection with Transformers," *European Conf. on Computer Vision (ECCV)*, 2020.DETR redefined object detection by using a transformer encoder-decoder framework. It eliminates post-processing steps and enables global reasoning—critical for modeling anatomical structures in medical imaging.
3. S. Naeem, M.U.G. Khan, M.A. Khan, et al., "Advances in Medical Image Analysis with Vision Transformers: A Comprehensive Review," *Computerized Medical Imaging and Graphics*, vol. 105, 2024.A large-scale review demonstrating that ViTs outperform CNNs across segmentation, classification, and detection tasks in medical imaging. Highlights effectiveness on small/low-res organs and average 0.8–2.5s inference time.
4. X. Zhu, W. Su, L. Lu, et al., "Deformable DETR: Deformable Transformers for End-to-End Object Detection," *International Conference on Learning Representations (ICLR)*, 2021.Introduced deformable attention to enhance DETR's efficiency and small object detection. The model achieved similar accuracy with 10x fewer epochs, ideal for organs of varying sizes in medical contexts.

5. Y. Wang, Z. Xu, X. Zhang, et al., "Anchor DETR: Query Design for Transformer-Based Detector," *IEEE Transactions on Image Processing*, vol. 31, pp. 1234–1246, 2022.Enhances DETR with anchor-based queries using positional priors. Achieved 44.2% AP on COCO and improved convergence speed, making it suitable for organ detection guided by anatomical positions.
6. Z. Liu, Y. Lin, Y. Cao, et al., "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012–10022, 2021. Swin Transformer introduces local windowed attention with global context shifts, ideal for high-resolution medical images. Its hierarchical design balances detail preservation with scalability.
7. D. Meng, X. Chen, Z. Fan, et al., "Conditional DETR for Fast Training Convergence," *IEEE/CVF International Conference on Computer Vision*, 2021.This model decouples spatial and content attention, improving DETR's training speed 10x and enhancing detection of small objects, which is vital for identifying minor anatomical features.
8. Touvron, M. Cord, M. Douze, et al., "Training Data-Efficient Image Transformers & Distillation Through Attention," *International Conference on Machine Learning (ICML)*, 2021.DeiT introduces knowledge distillation and data augmentation to train ViTs on smaller datasets. Particularly valuable for medical imaging where annotated data is often limited.
9. Lakshmi Narasimha Raju Mudunuri, Pronaya Bhattacharya, "Ethical Considerations Balancing Emotion and Autonomy in AI Systems," in Humanizing Technology With Emotional Intelligence, IGI Global, USA, pp. 443-456, 2025.
10. E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021. Explores XAI methods in healthcare, emphasizing transformer attention maps as natural explainability tools. Critical for clinical acceptance of AI-driven organ detection systems.
11. E.J. Topol, "High-Performance Medicine: The Convergence of Human and Artificial Intelligence," *Nature Medicine*, vol. 25, pp. 44–56, 2019.Discusses AI's integration in healthcare and the need for human-AI collaboration. Reinforces the importance of explainable, assistive tools in diagnostics rather than full automation.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details