Journal Homepage: -www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI:10.21474/IJAR01/21300
DOI URL:http://dx.doi.org/10.21474/IJAR01/21300

*RESEARCH ARTICLE*

**The National Conference on Innovative Trends in Modern Business Environment (ITMBE 2025) , DPG Institute of Technology and Management (DPGITM) Gurugram**

# TRANSFORMER SVS.CNN SIN MEDICAL IMAGING: A COMPARATIVE REVIEW

**Preeti Sharma[1], Poonam[1] and Alka yadav[2]**
1. Assistant Prof.,CAD, DPGITM, Gurugram,
2. Assistant Prof., CSE Department, GITM, Farrukh Nagar

## *Manuscript Info*

***Key words:-***
CNNs, Transformers , Deep Learning.

## *Abstract*

Deep learning has profoundly changed medical image analysis, with Convolutional Neural Networks (CNNs) serving as the conventio nal benchmark for tasks such as classification, segmentation, and detection. Lately, Transformers—initially created for natural language processing—have demonstrated significant achievements in computer vision and are currently being examined for medical imaging thanks to their capacity to grasp global context via self- attention methods. This review offers an extensive comparison between CNNs and Transformer s inmedical imaging,emphasizing their structural variations,advantages, and drawbacks.CNNsare proficient in local feature extraction and work well with small datasets, but frequently have difficulty in detecting longrange dependencies .Conversely, Transformers excel at capturing g lobal relationships, but they necessitate extensive datasets and significa nt computational power. Hybrid models that merge both architectur esa real so examined, presenting a promising gave nun to exploit their complementary advantages. This review seeks to assist researchers in choosing or creating suitable deep learning architectures for different medical imaging uses, concentrating on enhancing diagnostic precision and clinical significance.



## Introduction:-

The advent of deep learning has revolutionized the field of medical imaging, with Convolutional Neural Networks (CNN s) and Transformers emerging gas two dominant architectures. CNNs have long been the go-to choice for image analysis due to their ability to efficiently model local pixel interactions and their effectiveness in small scale dataset [1].However , the introduction of Transformers,whichexcelincapturingglobalrelationshipsthroughself attention mechanisms,Has prom pted are evaluation of the irrolein medica limaging [2][4].This article aim stoprovidea comprehensive comparison of these two architectures, highlighting their strengths, weaknesses, and potential future directions.

**Understanding of CNN**

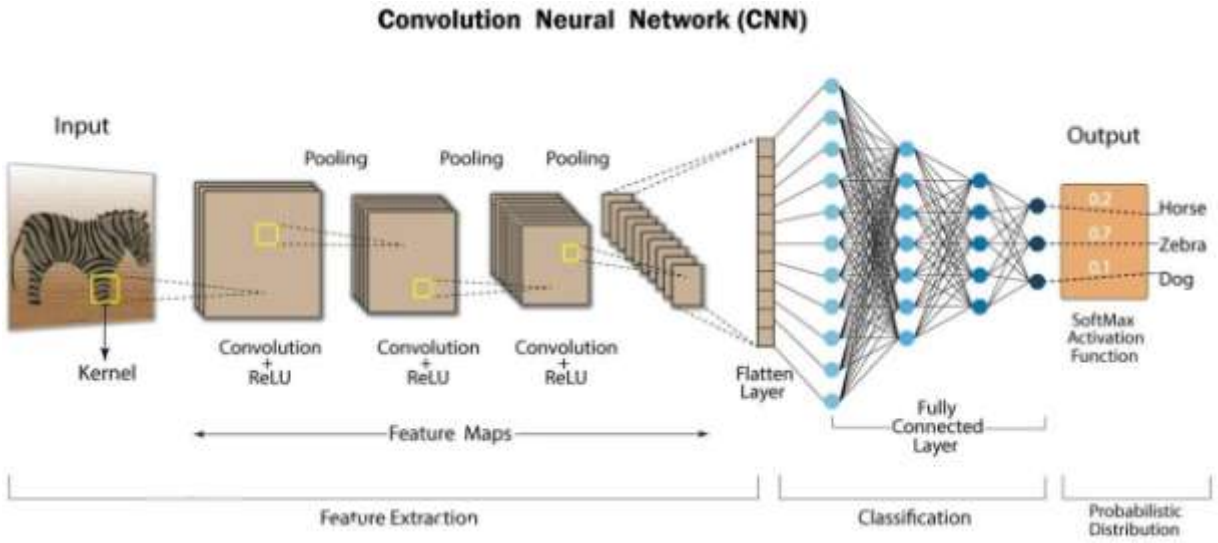


Fig.1

**Convolutional Neural Networks (CNNs) have served as the foundation for image processing tasks** for many years. They thrive at identifying local patterns using convolutional layers, which makes them especially suited for tasks like image classification, object detection, and segmentation. CNNsemployalayeredarchitecture,withinitia llayers identifying basic features (such sedges), while later layers recognize more intricate patterns (like shapes and objects) shown in fig.1. Nevertheless, CNNs face restrictions, especially in grasping long-range relationships because of their localized receptive fields. This may impede performance in activities that necessitate a comprehensive grasp of the input data, like video analysis or intricate scene comprehension**.**

**Advantages of CNNs**

1. **LocalFeatureExtraction**:CNN sare adeptat identifying local patterns,which is crucial for tasks like tumor detection and segmentation [1].
2. **DataEfficiency**:They can achieve competitive performan ceeven with smaller datasets, making them suitable for medical applications where annotated data is often scarce [11].
3. **Established Framework**: The extensive research and development surrounding CNNs have led to aplethora of pre-trained models and optimization techniques,facilitating their adoptionin clinical settings[1][6].

**Disadvantages of CNNs**

1. **LimitedGlobalContext**: CNNs struggle to model global relationships,which can becritical for understanding complex medical images [6][11].
2. **ComputationalComplexity**: Asthesize of the inputdata increases, CNNs can become computationally expensive, particularly in 3D imaging scenarios [6].

**Overview of Transformers**

Transformers, originally created for natural language processing, have become popular in computer vision because of capability to capture long range dependencies via self attention mechanisms. In contrast toCNNs, Transformers a nalyze the whole input at once, enabling themtoeffectivelyunderstandrelationshipsbetweenfar apartelements.This abi lity has resulted in their use across different fields,such as image classification an dvideo analysis.Recent research has shown that Transformers excelat managing long sequence time series inputs and forecasting tasks,surpassing co nventional CNNs in certain situations [2]. For example, the Tightly Coupled Convolutional Transformer (TCCT) model merges the advantages of CNN sand Transformers, improving efficiency and locality in forecasting time series [2].

**Advantages of Transformers**

1. **GlobalContextUnderstanding**: Transformers excelat capturinglong-rangedependencies, which can enhance the understanding of complex medical images [2][4].
2. **Scalability**:They can be scaled to handle large datasets,making them suitable for applications involving extensive medical imaging data [3][4].
3. **Robustness**: Transformers have shown resilience to input distortions, which is beneficial in clinical settings where image quality can vary [11].

**Disadvantages of Transformers**

1. **DataRequirements**: Transformers typically require large amount so fan notated data fore effective training, which can be a limitation in medical imaging [3][11].
2. **ComputationalIntensity**:The attention mechanisms in Transformers can lead to high computational costs, making them less efficient than CNNs in certain scenarios [6][11].

**Relative Performance**

Although CNNs have been the preferred architecture for image oriented tasks, Transformers have demonstrate enco uraging outcomes in multiple applications.For instance,in EEG classification, models based on Transformers have s urpassed conventional CNNs,highlightting their capability
In identifying time series signals [1]. In the same way, in the field of medical image segmentation, a hybrid method that merges CNNs with Transformers has demonstrated better results than employing each architecture independently [3].In remote sensing, the combination of CNNs and Transformers has resulted in notable progress in change detection tasks. The Asymmetric Cross-attention Hierarchical Network (ACAHNet) integrates both architectures to improve feature extraction and interaction, leading to better performance and lower computational expenses [6].

**Applications and Hybrid Models**

The integration of CNNs and Transformers has resulted in hybrid models that utilize the advantages of both frameworks. For example, the MSCANet, a CNN-Transformer network utilizing multiscale context aggregation, successfully captures hierarchical features and encodes long-distance contextual information [4]. In a similar vein, the Swin Transformer has been incorporated into CNN-oriented architectures for semantic segmentation, showing enhanced precision in remote sensing tasks [5]. Additionally, hybrid models are utilized in medical imaging, with the integration of CNNs and Transformers demonstrating potential in identifying conditions such as Alzheimer's disease [7]. The Pyramid Vision Transformer (PVT) has proven highly effective in deriving local and global features from MRI data, attaining strong accuracy in classification tasks [7].

**Obstacles and Prospective Paths**

Although Transformers offer benefits, they present challenges, especially regarding computational resources. Transformers usually need greater data and computational resources than CNNs, which can limit their accessibility for specific uses [8]. Moreover, although CNNs are simpler to optimize, Transformers may present more complexity because of their structure and training needs [9]. Future studies might aim to enhance the efficiency of Transformers, rendering them better suited for real-time applications. Furthermore, investigating new hybrid architectures that combine CNNs and Transformers more effectively may result in advancements across multiple areas, such as autonomous driving, medical imaging, and video analysis.

## Conclusion

To sum up, CNNs and Transformers each possess distinct advantages and disadvantages. Although CNNs are proficient at extracting local features and are simpler to optimize, Transformers surpass them in their ability to capture long-range dependencies and overall context. The continuous investigation of hybrid models that merge the advantages of both approaches shows significant potential for enhancing deep learning and its uses in different areas. As investigations progress, the field of deep learning architectures will probably transform, resulting in more efficient and effective approaches for intricate tasks.

## References

[1] MichaelT.McCann;KyongHwanJin;MichaelUnser;"ConvolutionalNeuralNetworksforInverse Problems in Imaging: A Review", IEEE SIGNAL PROCESSING MAGAZINE,2017.
[2] JunLi; Junyu Chen;YuchengTang;CeWang; BennettA. Landman;S. KevinZhou; "Transforming Medical Imaging with Transformers? A Comparative Review of KeyProperties, Current Progresses, and Future Perspectives",ARXIV-CS.CV,2022.
[3] Junfei Xiao;Yutong Bai ;Alan Yuille; ZongweiZhou;"Delving Into Masked Autoencodersfor Multi-Label Thorax Disease Classification",ARXIV-CS.CV,2022.
[4] Emerald U. Henry;OnyekaEmebob;ConradAsotieOmonhinmin;"Vision Transformers inMedical Imaging: A Review",ARXIV-CS.CV,2022.
[5] Fahad Shamshad; SalmanKhan; Syed Waqas Zamir; Muhammad Haris Khan;MunawarHayat;FahadShahbazKhan;HuazhuFu;"TransformersinMedicalImaging:ASurvey",MEDICAL IMAGE ANALYSIS,2023.
[6] Giorgos Papanastasiou; NikolaosDikaios; JiahaoHuang; ChengjiaWang; Guang Yang;"IsAttention All You Need in Medical Image Analysis? A Review", IEEE JOURNAL OFBIOMEDICAL AND HEALTH INFORMATICS,2024.
[7] Moein Heidari;Sina Ghorbani Kolahi;SanazKarimijafarbigloo;BobbyAzad;AfshinBozorgpour;SoheilaHatami;RezaAzad;AliDiba;UlasBagci; Dorit Merhof; Ilker Hacihaliloglu;"Computation-Efficient Era: A Comprehensive Survey of State Space Models in Medical Image Analysis",ARXIV-EESS.IV,2024.
[8] J. W. Kim;A. U. Khan; I. Banerjee; "Systematic Review of Hybrid Vision Transformer Architectures for Radiological Image Analysis",MED.RADIOLOGY-AND-IMAGING,2024.
[9] Satoshi Takahashi; Yusuke Sakaguchi; Nobuji Kouno;Ken Takasawa;KenichiIshizu;YuAkagi;RinaAoyama;NaokiTeraya;AminaBolatkan;NorioShinkai;HidenoriMachino;KazumaKobayashi;KenAsada; Masaaki Komatsu; Syuzo Kaneko; Masashi Sugiyama; Ryuji Hamamoto; "Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review", JOURNAL OF MEDICALSYSTEMS,2024.
[10] PooyaMohammadiKazaj;GiovanniBaj;YazdanSalimi;AnselmW.Stark;WaldoValenzuela;GeorgeCM.Siontis;HabibZaidi;MauricioReyes;ChristophGraeni; IsaacShiri; "From Claims to Evidence: A Unified Framework and Critical Analysis of CNN Vs.Transformer Vs. Mamba in Medical Image Segmentation",ARXIV-EESS.IV,2025.
[11] Wikipedia: Vision transformer.
[12] Jiayao Sun; Jin Xie; Huihui Zhou;"EEG Classification with Transformer-BasedModels",2021IEEE3RDGLOBALCONFERENCEONLIFESCIENCESAND...,2021.
[13] LiShen;YangzhuWang;"TCCT:Tightly-CoupledConvolutionalTransformeronTimeSeries Forecasting",ARXIV-CS.LG,2021.
[14] Xiangde Luo; Minhao Hu; Tao Song; Guotai Wang; Shaoting Zhang;"Semi-SupervisedMedicalImageSegmentationViaCrossTeachingBetweenCNNandTransformer",ARXIV-EESS.IV,2021.
[15] MengxiLiu;ZhuoqunChai;HaojunDeng;RongLiu;"ACNN-TransformerNetworkWithMultiscale Context Aggregation for Fine-Grained Cropland Change Detection",IEEEJOURNAL OF SELECTED TOPICS IN APPLIED EARTH ...,2022.
[16] Xingwei He; Yong Zhou; Jiaqi Zhao; Di Zhang; Rui Yao; Yang Xue;"Swin TransformerEmbeddingUNetforRemoteSensingImageSemanticSegmentation",IEEETRANSACTIONSON GEOSCIENCE AND REMOTE SENSING,2022.
[17] Xiaofeng Zhang; Shuli Cheng; Liejun Wang; HaojinLi;"Asymmetric Cross-AttentionHierarchicalNetworkBasedonCNNandTransformerforBitemporalRemoteSensingImagesChange Detection",IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTESENSING,2023.
[18] Vision Transformersvs. Convolutional Neural Networks | by Fahim..