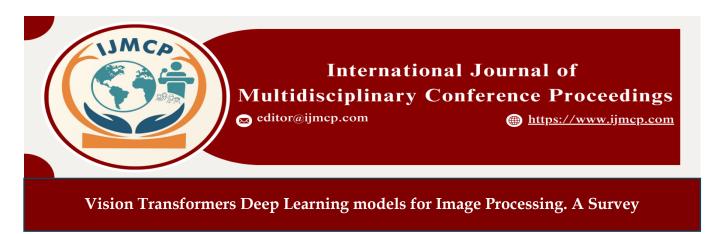
Sana Cheema, Ali Daud, Tariq Alsahfi, Hikmat Ullah Khan & Kainat Naeem



1st Sana Cheema, 2nd Ali Daud, 3rd Tariq Alsahfi, 4th Hikmat Ullah Khan & 5th Kainat Naeem

1^{st,} 2nd, 3rd, 4th, 5th The Islamia University Bahawalpur, Pakistan

ABSTRACT
Transformers have emerged as a novel and transformative technology in deep learning, providing an alternative to image processing which was historically based on Convolutional Neural Network. The present study demonstrates the adaptability of Vision Transformers from natural Language processing to a variety of image analysis applications. VIT incorporates elements from the architecture, such as the self-attention mechanism, that permits the model to capture global dependencies in the images. The performance of ViT is also evaluated compared to CNN. The researchers emphasize the advantages that ViT have concerning the processing of complex visual information and the disadvantages they face because of their computational and data requirements. The paper also briefly considers the wide spectrum of ViT applications in medical imaging, object detection, and others. Thus, the paper discusses the current trends and expected developments and states that Vision Transformers
may be the key to the intensification of image processing technologies.
 sanaacheema34@gmail.com 2(1) 19 Cheema, S., Daud, A., Alsahfi, T., Khan, H. U., & Naeem, K. (2025). Vision Transformers Deep Learning models for Image Processing. A Survey. Proceedings of the 1st International Conference on Innovation and Sustainability in Management and Social Sciences, *International Journal of Multidisciplinary Conference Proceedings, 2(1).