

Application of a hybrid Vision Transformer and LSTM model in predicting circumferential contraction timing from cardiovascular image sequences.

Ha Quoc Trung

Abstract: The thesis proposes a smart medical solution combining Vision Transformer (ViT) and LSTM to analyze echocardiography image sequences. ViT handles the extraction of detailed spatial features from frames, while LSTM learns temporal dependencies to accurately predict the contraction timing of the heart muscle. This method helps physicians evaluate left ventricular function and diagnose cardiovascular diseases earlier in a more automated and objective manner. Research results show high correlation compared to traditional clinical diagnostic methods.

Keywords: Echocardiography, Vision Transformer (ViT), LSTM, Image sequence analysis

References

- [1] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” arXiv preprint, 2021. [2] D. Ouyang et al., “Video-based AI for beat-to-beat assessment of cardiac function,” *Nature*, vol. 580, pp. 252–256, 2020. [3] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.