

# Comparison of Models for Automatic Description of Medical Images

*Thursday 24 April 2025 10:45 (30 minutes)*

Automated medical report generation from chest X-ray images is a critical area of research in medical AI, aiming to enhance diagnostic accuracy, reduce radiologists' workload, and improve patient care. The process involves analyzing medical images and translating visual findings into structured, clinically relevant textual reports. Traditional methods rely on human expertise, which is time-consuming and prone to variability, motivating the development of deep learning-based solutions that leverage vision-language models to automate this task. This project explores and compares state-of-the-art deep learning architectures for medical report generation, evaluating their capabilities in image encoding and text generation. The study considers convolutional neural networks (CNNs) such as ResNet, vision transformers (ViTs) like SwinTransformer, and state space models (SSMs) such as Mamba for extracting visual features. The text generation stage utilizes recurrent neural networks (RNNs) such as LSTMs and GRUs, as well as transformer-based architectures such as BioClinicalBERT, LLaMA-2, and GPT-style decoders. The models being evaluated include BioViL-T, R2Gen, MedCLIP, PLIP, CheXbert, and MambaXray-VL, trained and tested on datasets such as IU X-Ray and CheXpert. This study aims to systematically assess different architectural approaches, training methodologies, and dataset utilization strategies to provide insights into their advantages and limitations in generating clinically meaningful radiology reports.

**Author:** Mr URBAŃSKI, Jakub (Warsaw University of Technology)

**Presenter:** Mr URBAŃSKI, Jakub (Warsaw University of Technology)

**Session Classification:** Session B (Poster)