

# Explainability of Convolutional Neural Network: Overview of Methods

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

Understanding and interpreting the decisions made by deep learning models has become an essential area of research in artificial intelligence. Convolutional neural networks (CNNs), despite their high performance in various tasks, often function as “black boxes,” making it challenging to explain their predictions. This study focuses on applying and evaluating different explainability techniques to CNN models to gain more insight into their decision-making processes. Using multiple approaches, our aim was to assess the effectiveness and reliability of these methods in improving the transparency and interpretability of neural networks.

**Authors:** KRUSZYŃSKI, Jan (PW); Mr SONBUL, Karim (PW); Mr MARKOWICZ, Magdalena (PW); Mr PODGÓRSKI, Piotr (PW)

**Co-author:** SIWEK, Krzysztof

**Presenter:** Mr MARKOWICZ, Magdalena (PW)

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