



### Joint European Conference on Machine Learning and Knowledge Discovery in Databases

ECML PKDD 2021: **Machine Learning and Knowledge Discovery in Databases. Research Track** pp 633–649

## Causal Explanation of Convolutional Neural Networks

[Hichem Debbi](#) 

Conference paper | [First Online: 10 September 2021](#)

**1051** Accesses

Part of the [Lecture Notes in Computer Science](#) book series (LNAI, volume 12976)

### Abstract

---

In this paper we introduce an explanation technique for Convolutional Neural Networks (CNNs) based on the theory of causality by Halpern and Pearl [12]. The causal explanation technique (CexCNN) is based on measuring the filter importance to a CNN decision, which is measured through counterfactual reasoning. In addition, we employ extended definitions of causality, which are responsibility and blame to weight the importance of such filters and project