

Title: The efficacy of machine learning models in lung cancer risk prediction with explainability

Author Names: Refat Khan Pathan, Israt Jahan Shorna, Md. Sayem Hossain, Mayeen Uddin Khandaker, Huda I. Almohammed, Zuhal Y. Hamd

Abstract: Among many types of cancers, to date, lung cancer remains one of the deadliest cancers around the world. Many researchers, scientists, doctors, and people from other fields continuously contribute to this subject regarding early prediction and diagnosis. One of the significant problems in prediction is the black-box nature of machine learning models. Though the detection rate is comparatively satisfactory, people have yet to learn how a model came to that decision, causing trust issues among patients and healthcare workers. This work uses multiple machine learning models on a numerical dataset of lung cancer-relevant parameters and compares performance and accuracy. After comparison, each model has been explained using different methods. The main contribution of this research is to give logical explanations of why the model reached a particular decision to achieve trust. This research has also been compared with a previous study that worked with a similar dataset and took expert opinions regarding their proposed model. We also showed that our research achieved better results than their proposed model and specialist opinion using hyperparameter tuning, having an improved accuracy of almost 100% in all four models.

Keywords: Lung cancer prediction Machine learning models Explainable AI (XAI) Model interpretability Hyperparameter tuning Performance comparison

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0305035>