

# Gene Selection Algorithms: Literature review

Borislava Vrigazova<sup>1</sup>

<sup>1</sup>Sofia University, bul. Tsarigradsko Shosse, vrigazova@uni-sofia.bg

## Abstract

Keywords: gene selection, data mining, cancer

Machine learning algorithms have made the topic of gene selection in cancer datasets very important. Gene selection algorithms have become the subject of much research. They have important implications for identifying genes that can participate in developing cancer [1]. The selected genes [2] can further be tested in medical laboratory to understand and prevent cancer more effectively. The aim of this research is to summarize the most often used gene selection algorithms leading to cancer. Such type of summary is the first step of improving existing algorithms and devising new ones. We summarize the most often used techniques for gene selection in cancer datasets and the important genes for cancer development based on existing studies. We also try to clarify what combinations of genes are responsible for cancer development according to medical studies and how they correspond to the results from machine learning models. This approach would help us identify gaps in existing algorithms that can be improved and gather ideas for novel approach towards gene selection in cancer datasets.

## References

1. Ting, F., Tan, Y., & Sim, K. (2019). Convolutional neural network improvement for breast cancer classification. *Expert Systems with Applications*, 120, 103–115.
2. Singh, B. (2019). Determining relevant biomarkers for prediction of breast cancer using anthropometric and clinical features: A comparative investigation in machine learning paradigm. *Biocybernetics and Biomedical Engineering*, 39(2), 393–409