Comprehensive study on importance of feature selection methods to predict cancer tumor types

Gopi Battineni (✉ gopi.battineni@unicam.it)

University of Camerino <https://orcid.org/0000-0003-0603-2356>

Francesco Amenta

Università degli Studi di Camerino Scuola di Scienze del Farmaco e dei Prodotti della Salute

Nalini Chintalapudi

Jawaharlal Nehru Technological University - Anantapur

Research article

Keywords:

Posted Date: September 5th, 2019

DOI: <https://doi.org/10.21203/rs.2.11833/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Predictions on cancer development were getting better through Machine Learning (ML) methods compared to pathologists. This study is aim to develop a ML model based feature selection methodology to get accurately predict the breast cancers Methods: We consider database of 569 breast cancer cases and diagnosed with 212 malignant and 357 benign and this study adopt with logistic regression (LR) ML type that had done with total feature set and selective feature set. Results: Outcomes produces cancer prediction accuracy 90.86% with total features and 93.84% with selective features. In addition, we validated the results with Area under Curve (AUC) for the selective feature set Receiver Optimistic curve (ROC) curve is greater (99.8%) than AUC for the total feature set ROC (96.2%). In precise, LR model did accurate tumor classification with selective features. Conclusions: It is very important to address issues of cancer tumor classifications by ML and it will help pathologists to provide preventive care for the patients. However, it is important to involve other features of cancer in order to understand the further causes and decision-making process that should be help in much accurately classified.