IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 9, May 2025

Impact Factor: 7.67

An Ensemble Model for Multi-Label Classification of Biomedical Big Data in Breast Cancer Research

Dr. Prem Kumar Chandrakar
Department of Computer Science,
Mahant Laxminarayan Das College, Raipur (C.G.) India.
premchandrakar@gmail.com

Abstract: *The exponential increase in high-throughput biomedical data has introduced substantial challenges in processing complex, multi-label, and high-dimensional datasets—especially within the context of breast cancer research. Conventional single-label classification techniques often fail to account for the intricate associations among multiple clinical outcomes. This research evaluates the effectiveness of ensemble-based multi-label classification (MLC) techniques on the TCGA-BRCA dataset, which integrates both genomic and clinical information. We analyze the performance of three well-known MLC algorithms—Binary Relevance (BR), Classifier Chains (CC), and Random k-Labelsets (RAkEL)—in combination with base classifiers such as Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Machine (GBM). Our experimental results show that the ensemble model consistently surpasses individual approaches in metrics like Hamming Loss, Exact Match Ratio, and both Micro and Macro F1-scores. These findings underscore the strength and reliability of ensemble learning in biomedical data contexts, suggesting strong potential for its use in clinical decision support and personalized oncology.*

Keywords: multi-label classification, ensemble learning, breast cancer, biomedical big data, TCGA-BRCA, machine learning, classification performance

DOI: 10.48175/IJARSCT-27032