

Cancer Biomarker Discovery Using High Performance Algorithm

Maryam Razmjouei^a, Hamid Reza Hamidi^{*a}, Esmail Ebrahimi^b

^a Computer Engineering Department, Imam-Khomeini International University, 34148-96818, Qazvin, Iran

^b Department of Biotechnology, University of Shiraz, Shiraz, Iran

* hamidreza.hamidi@eng.ikiu.ac.ir

Abstract: Biomarker has an important role for early diagnosis and also it can predict progression disease and provide monitoring treatment. Biomarker is a structure in the body that has biological changes due to a specific disease. Analyze nucleic acids, protein and metabolism may provide cancer biomarkers. Biomarkers discovery and study of their status is possible using bioinformatics tools. However, Due to high-dimensional data and several types of them along with lack of specific method for biomarker discovery, choosing the biomarkers that have diagnostic and predictive information is not easy. In this paper, we introduce bioinformatic tools for biomarker discovery such as datasets and computational tools then we focus on feature selection algorithms as a house for biomarker discovery. Using feature selection algorithms in addition to the given a subset of important and relevant features, reduce data dimensional and increase accuracy. In this way we introduce the best feature selection methods and classification algorithms used for this purpose and discuss computational challenges such as performance and stability. In this research we select breast cancer and provide data from The Cancer Genome Atlas (TCGA) data portal that include 1800 patients and more than 2500 attributes. At the end of this paper we introduce parallel computing as a solution for high-dimensional data analysis.

Keywords: Biomarker, Feature Selection algorithms, Parallel Computing, Breast Cancer, TCGA.

References

- [1]. Mahmood A. Mahdavi, "Bioinformatics – Trends and Methodologies", October, 2011.
- [2]. Li Wu, Xiaogang Qu, "Cancer biomarker detection: recent achievements and challenges", The Royal Society of Chemistry 2015.
- [3]. Jason E. McDermott, Jing Wang, Hugh Mitchell, "Challenges in Biomarker Discovery: Combining Expert Insights with Statistical Analysis of Complex Omics Data", Bobbie-Jo Webb-Robertson, Expert Opin Med Diagn. Author manuscript; available in PMC 2013 Jan 18.
- [4]. Malik Yousef, Naim Najami, "Computational Approaches for Biomarker Discovery, The Institute of Applied Research", The Galilee Society, Shefa Amr, Israel, Computer Science, The College of Sakhnin, Sakhnin, Israel, Journal of Intelligent Learning Systems and Applications, 2014, 6, 153-161.
- [5]. Tomczak, K., Czerwinska, P., and Wiznerowicz, "The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemporary Oncology", (2015), 19(1A): A68-A77.
- [6]. Martinez-Ledesma, Emmanuel, Verhaak, Roeland G.W., Treviño, Victor, "Identification of a multi-cancer gene expression biomarker for cancer clinical outcomes using a network-based algorithm", Scientific Reports 5, Article number: 11966 (2015).
- [7]. Aguirre-Gamboa R, Gomez-Rueda H, Martínez-Ledesma E, Martínez-Torteya A, Chacolla-Huaranga R, et al. "SurvExpress: An Online Biomarker Validation Tool and Database for Cancer Gene Expression Data Using Survival Analysis". 2013, PLOS ONE 8(9): e74250
- [8]. Phan, John H., Andrew N. Young, and May D. Wang. "omniBiomarker: A Web-Based Application for Knowledge-Driven Biomarker Identification." IEEE transactions on bio-medical engineering 60.12 (2013): 3364–3367. PMC. Web. 30 June 2017.