Breast Cancer Risk Prediction using Statistical Modeling and Predictive Machine Learning Analytics Modeling

Anonymous Submission[1] **Anonymous Submission[2]** **Anonymous Submission[3]**

Abstract

Breast cancer claimed the lives of 74,072 people in Africa in 2018, with an estimated 168,690 cases [2]. The age-standardized incidence (World) rate was 37.9 per 100,000 ranging from 6.9 per 100,000 in Gambia to 69.6 per 100,000 in Mauritius [2]. Breast cancer in Moroccan women increased from 35.0 to 39.0 per 100,000 women between 2004 and 2008, with an annual increase of 2.85 %. The highest incidence rates were registered in the age groups of 45-49, 50-54 and 55-59 years respectively [1]. Sixty-nine percent of female breast cancer cases were diagnosed at stages II and III and in 2012, the estimated number of women who died of breast cancer in Morocco was 2,878. The crude, age-standardized (World) mortality rates were 17.3 and 18.0 per 100,000, respectively [1]. Breast cancer accounts for 27.7% of total cancer cases in African nations, followed by cervical cancer, which accounts for 19.6% of total cases. When taken as a whole, this is the most prevalent in African women [3].

Breast cancer is the most common malignancy among women, accounting for nearly 1 in 3 cancers diagnosed among women in the World, and it is the second leading cause of cancer death among women [4]. Breast cancer occurs because of abnormal growth of cells in the breast tissue, commonly referred to as a tumor. A tumor does not mean cancer - tumors can be benign (not cancerous), or malignant (cancerous). Tests such as MRI, mammogram, ultrasound, and biopsy are commonly used to diagnose breast cancer, which are not automated methods of diagnosing breast cancer. Automated classification of breast cancer mammogram dataset should help as an initiation to build a more efficient and objective screening in healthcare. While past research has established a few artificial intelligence (AI) algorithms for breast cancer classification, there are several other findings that must be considered, and complete and accurate models of these major findings have yet to be published. In this project, we aim to apply statistics predictive models and machine-learning analytics for detecting breast cancer on numerical dataset for important features necessary for detection and diagnosis.

References

1. Khalis, M., El Rhazi, K., Charaka, H., Chajès, V., Rinaldi, S., Nejjari, C., ... & Charbotel, B. (2016). Female breast cancer incidence and mortality in Morocco: comparison with other countries. Asian Pacific Journal of Cancer Prevention: APJCP, 17(12), 5211. APJCP Journal

2. Sharma, R. (2021). Breast cancer burden in Africa: evidence from GLOBOCAN 2018. Journal of Public Health, 43(4), 763-771. https://pubmed.ncbi.nlm.nih.gov/32657321/

3. Bahnassy, A. A., Abdellateif, M. S., & Zekri, A. R. N. (2020). Cancer in Africa: is it a genetic or environmental health problem?. Frontiers in Oncology, 10, 604214. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7781064/

4. American Cancer Society (2022). Key Statistics for Breast Cancer: How common is breast cancer?. Retrieved from: https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html#:~:text=It%20is%20about%2030%25%20

Appendix:

Google Research Colab Code Implementation

Our source code is available on request because the research is still on going and still in early stage.