

Machine Learning Analysis of a Chilean Breast Cancer Registry

Francisco Acevedo^{1,2}, Leonardo Causa³, Sebastián Bravo³, Pablo García³, Ricardo Cuevas³, Maria Loreto Bravo¹, Carla Avellaira⁴, Sabrina Muñoz², Militza Petric², Raúl Martínez², Constanza Guerra², Marisel Navarro², Carla Taramasco⁵ and Cesar Sanchez^{1*}



¹Department of Hematology-Oncology, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile

²Complejo Asistencial y Hospital Dr. Sotero del Rio, Santiago, Chile

³Research and Development Department, B-Trials SpA, Facultad de Ingeniería y Negocios, Universidad de las Américas, Santiago, Chile

⁴Real World Evidence, Medical Affairs Department, Roche Chile, Santiago, Chile

⁵Escuela de Ingeniería Civil Informática, Universidad de Valparaíso, Centro Nacional de Sistemas de Información en Salud (CENS), Valparaíso, Chile

*Corresponding author: Cesar Sánchez, Department of Hematology-Oncology Facultad de Medicina Pontificia Universidad Católica de Chile Diagonal Paraguay 362, 6th FL RM608 Postal Code: 8330077, Santiago, Chile

ARTICLE INFO

Received: 📅 July 22, 2021

Published: 📅 August 02, 2021

Citation: Francisco Acevedo, Leonardo Causa, Sebastián Bravo, Pablo García, Cesar Sanchez, et al., Machine Learning Analysis of a Chilean Breast Cancer Registry. Biomed J Sci & Tech Res 37(4)-2021. BJSTR. MS.ID.006037.

ABSTRACT

In recent years, artificial intelligence (AI) and machine learning (a form of AI) have offered valuable tools for medicine by applying and training algorithms in order to make predictions. Herein, we applied a machine learning algorithm to analyze data from a >20 year breast cancer (BC) registry elaborated in two Chilean health institutions (a public hospital and a private center) that includes a total of 4838 patients and their basic clinical-pathological characteristics. Preliminary results suggest that this cohort of patients can be subdivided into five clusters according to key variables that also correlate with overall survival and disease-free survival rates. To our knowledge this is the first Latin American report of its kind. Our laboratory is currently expanding these analyses.

Keywords: Breast Cancer; Machine Learning; Overall Survival; Disease-Free Survival

Purpose

As occurs in several countries, breast cancer (BC) is one of the leading causes of cancer related death among Chilean women [1]. Like other malignancies, breast neoplasms are characterized by their heterogeneity. This not only applies to clinical features of patients but also, to molecular, genetic and histologic characteristics [2]. Similarly, incidence rates and associated risk factors display a marked geographic variability [1]. To date, several studies have reported BC incidence and prevalence rates in both Europe and North America. These studies have also reported clinical-genetic characteristics and prognosis. In sharp contrast, South American

reports on these topics are scarce [3]. Indeed, only a few Latin American studies have included data on limited populations, these are mostly from Brazil and Mexico [4,5]. Unpublished data from our group suggest that differences in lifestyle along with a diverse racial background could explain particular characteristics observed in the Chilean population.

In recent decades, Artificial Intelligence (AI) has emerged as an innovative and valuable tool in medicine, providing assistance to achieve more accurate patient diagnoses and to support making medical decisions. Interestingly, certain studies demonstrate that

AI-algorithms can compete or even outperform clinicians in specific tasks [6]. In lay terms, AI-algorithms can be easily 'trained' by using sample data. Thus, algorithms "learn" to do their job much like doctors learn by attending medical school for years, making right decisions and sometimes mistakes. Within this context, Machine Learning (a form of AI) seeks to apply algorithms and build models based on training data in order to make predictions in a variety of applications including medicine [7]. In 1997 our institution started a longitudinal BC registry that included invasive disease cases. In recent years, our group has generated several publications focused on BC incidence, clinical characteristics of patients and clinical data based on these analyses [8-10]. Herein we report preliminary analyses on data applying machine learning to analyze our local BC patient registry.

Patients and Methods

This study was part of a collaborative effort between Hospital Sotero del Rio and Cancer Center at Pontificia Universidad Católica de Chile, the former a public hospital and the later a university cancer center, both at Santiago, Chile. We sought to determine relevant clusters of BC patients associated with clinical characteristics and survival that allow us to evaluate and propose patient-adapted therapeutic schemes. The K-medoids clustering algorithm was used

to define a patient profile based on demographic (sex, age, weight / height, cancer family history, comorbidities and BC risk factors) and clinical-pathological information (stage, BC subtype, surgery, type of systemic treatment). Once the groups were separated, survival rates were calculated using the Kaplan-Meier method. This analysis allows us to link patient profiles with the behavior of survival rates. Then, data analytics methods were applied to determine the most relevant variables for each of the clusters and their correlation with survival rates. Finally, we estimate the time evolution of the treatments carried out (trajectories). In this way, it is possible to describe treatment schemes for each of the defined clustering.

Results

Overall, a total of 4838 registered BC patients were included into our study. Our analyses divided patients into five clusters with marked differences in clinical characteristics and prognoses see Figure 1. The key variables that defined these clusters included: age at diagnosis, body mass index, family history of cancer (by a first-degree relative), comorbidities (mainly hypertension), compromised nodes, and BC relapse. Clusters were also associated with significant differences in overall and disease-free survival (Figure 2).

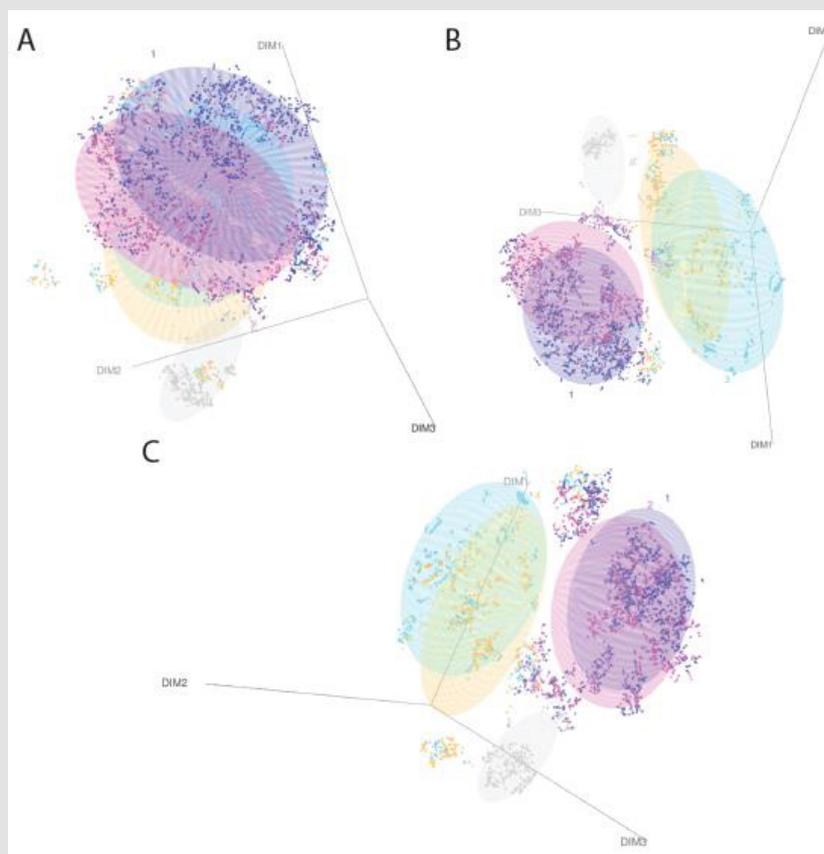


Figure 1: Graphical three-dimensional representation of the five clusters of patients generated by our model. Panels A, B and C show different angles.

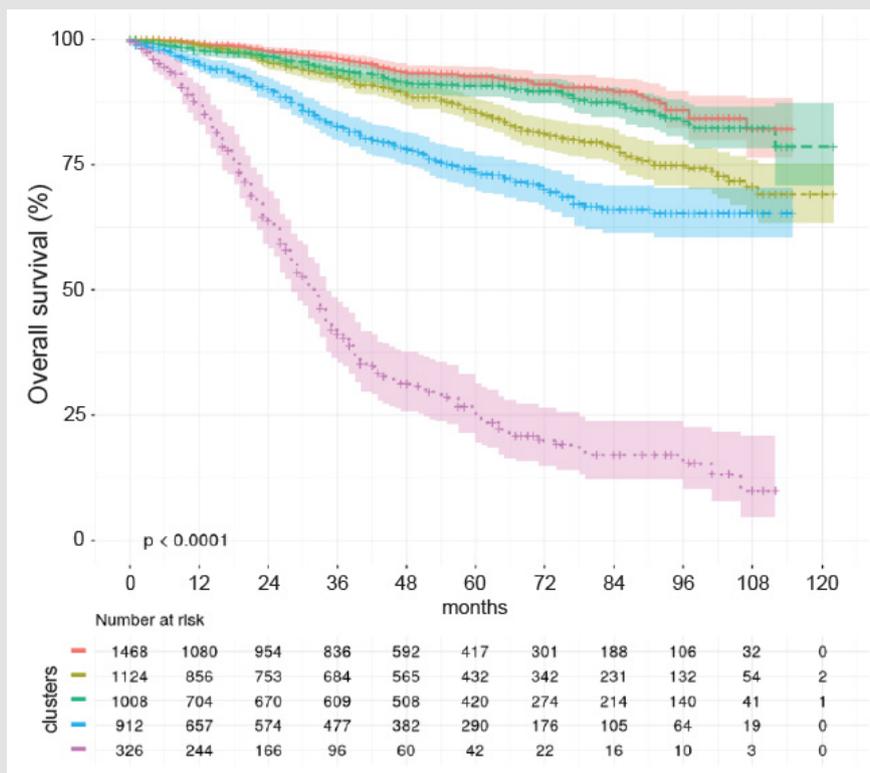


Figure 2: Overall survival for the identified clusters of patients.

Conclusion

To our knowledge, this is the first Latin American report applying a machine learning approach to analyze BC registry data, including clinical features and survival outcomes. Our findings confirm the capacity of machine learning to differentiate BC clusters with specific clinical and prognostic outcomes. Currently, we are validating this approach and expanding our database.

References

- (2021) Global cancer observatory.
- Reinert T, De Souza ABA, Parisotto Sartori G, Obst FM, Barrios CH (2021) Highlights of the 17th St. Gallen International Breast Cancer Conference 2021: customising local and systemic therapies. *Ecanermedalscience* 18: 15.
- Justo N, Wilking N, Jönsson B, Luciani S, Cazap E (2013) A Review of Breast Cancer Care and Outcomes in Latin America. *Oncologist* 18(3): 248-256.
- Goss PE, Lee BL, Badovinac-Crnjevic T, Strasser-Weippl K, Chavarri-Guerra Y, et al. (2013) Planning cancer control in Latin America and the Caribbean. *Lancet Oncol* 14(5): 391-436.
- Cazap E (2018) Breast Cancer in Latin America: A Map of the Disease in the Region. *Am Soc Clin Oncol Educ B* 38: 451-456.
- Liefaard MC, Lips EH, Wesseling J, Hylton NM, Lou B, et al. (2021) The Way of the Future: Personalizing Treatment Plans Through Technology. *Am Soc Clin Oncol Educ B* 41: 12-23.
- Leite ML, De Loiola Costa LS, Cunha VA, Kreniski V, De Oliveira Braga Filho M, et al. (2021) Artificial intelligence and the future of life sciences. *Drug Discov Today*.
- Maiz C, Silva F, Domínguez F, Galindo H, Camus M, et al. (2020) Mammography correlates to better survival rates in breast cancer patients: a 20-year experience in a University health institution. *Ecanermedalscience* 23: 14.
- Acevedo F, Petric M, Walbaum B, Robin J, Legorburu L, et al. (2021) Better overall survival in patients who achieve pathological complete response after neoadjuvant chemotherapy for breast cancer in a Chilean public hospital. *Ecanermedalscience* 11: 15.
- Walbaum B, Puschel K, Medina L, Merino T, Camus M, et al. (2021) Screen-detected breast cancer is associated with better prognosis and survival compared to self-detected/symptomatic cases in a Chilean cohort of female patients. *Breast Cancer Res Treat*.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2021.37.006037

Cesar Sanchez. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>