New Real Dataset Creation to Develop an Intelligent System for Predicting Chemotherapy Protocols

Case of Moroccan Breast Cancer Patients

Houda AIT BRAHIM[1], Mariam BENLLARCH[2], Nada BENHIMA[3],
Salah EL-HADAJ[4], Abdelmoutalib METRANE[5], Ghizlane BELBARAKA[6]
Computer and Systems Engineering Laboratory-Faculty of Science and Technology, Cadi Ayad University, Marrakech, Morocco[1, 2, 4, 5]
Science and Health Laboratory, Faculty of Medicine, Marrakech, Morocco[3, 6]

***Abstract*—Breast cancer is the most common cancer diagnosed in women. In developing countries, controlling this scourge is often problematic due to late diagnosis and the lack of medical and human resources. Automation and optimization of treatment is then needed to improve patient outcome. The use of medical datasets could, according to medical staff and pharmacists, assist them in clinical decision-making and would allow for better use of resources especially when limited. In our paper, a new real dataset was produced by collecting medical and personal data from 601 patients with breast cancer at the University Hospital Center (UHC) Mohammed VI of Marrakech. Data of women diagnosed with breast cancer from January 2018 at UHC were assessed. Most patients were 24-85 year-old, with an average age of 48.84 years. Patient age, performance status (PS), cancer stage and subtype, treatment patterns and correlations among the different variables were analyzed. The created dataset will help to determine the most appropriate treatment regimen depending on the individual characteristics of patients to allow for better use of limited resources.**

***Keywords*—Dataset; breast cancer; cancer stage; chemotherapeutic regimen; machine learning; prediction**

I. INTRODUCTION

According to World Health Organization (WHO) [1]:

- Breast cancer is the most common and deadliest malignancy in women.
- In 2020, around 2.3 million women were diagnosed with breast cancer and around 685,000 deaths were recorded worldwide due to this malignancy.
- It may strike pubescent women from all over the world at any time,
- The risk increases with age.

The frequency of breast cancer continues to increase [2].

The exact causes of breast cancer are unknown. They are complex, multifactorial and depend on personal characteristics [3]. The study [4] reported that environmental factors are most likely the primary reasons behind breast cancer rather than genetic factors. According to [5], the causes of breast cancer were believed to be either behavioral (e.g. physical inactivity and alcohol intake) or non-behavioral (age and genetic factors). However, the most common causes of breast cancer were found to be: (i) stress and worry, (ii) diet and eating habits, (iii) altered immunity, (iv) overwork and, (iv) poor previous medical care. Similarly, [3] found that stress and lifestyle-associated factors (e.g. diet, exercise and weight control) are the predominant causal agents of breast cancer.

Different therapies are used to treat breast cancer. This depends on cancer subtype and progression [6]. For example, surgery, with either lumpectomy (i.e. removing the tumor) or total mastectomy (i.e. removing the whole breast); radiotherapy, which includes external radiotherapy (i.e. high energy radiations are aimed at the breast tumor) and brachytherapy (i.e. a radioactive compound is injected into the breast); chemotherapy, through drug administration to prevent metastasis or reduce the size of an existing tumor; hormone therapy, by using either drugs, radiation or surgery to reduce female hormone activity and production; and targeted therapy coupled with immunotherapy, in which drugs are used to stimulate the immune system [7].

In Morocco, breast cancer represents around 35% of all cancers diagnosed in women, with the highest incidence rate observed in 45-59-year-old women [8] [9]. Like many other developing countries, breast cancer represents a major health concern in Morocco [10]. This is due not only to the lack of early diagnosis of this cancer, particularly in the countryside, but also to the lack of well-established treatment options for patients depending on their personal, metabolic and medical profiles. On the other hand, the acquisition of anticancer drugs is a serious problem for developing countries. Indeed, the cost of anticancer drugs used in different therapeutic regimes is extremely high and continues to rise [11]. Taking this into account along with the fact that not all Moroccan citizens and cancer treatments are covered by health insurance, and the lack of human and medical resources make it necessary to carefully manage finite resources and to optimize cancer treatment for each patient.

Today, developing countries are facing the challenge of improving patient outcomes and survival with limited access to advanced medical care [12]. Most of these countries can only provide conventional chemotherapeutic agents to their citizens [13]. Indeed, targeted therapies such as cyclin-dependent kinase inhibitors, anti-human epidermal growth factor receptor

2 (anti-HER2), tyrosine kinase inhibitors and bifunctional monoclonal antibodies (i.e. antibody–drug conjugates) are available in limited quantities and not covered by health insurance [14] [15]. Along this line, determining the optimal therapy depending on each patient specific characteristics would improve patient outcomes and satisfaction even when resources are limited [16]. In such context, the use of medical datasets would be of great interest to implement a convenient decision-making program to help physicians in identifying the most appropriate therapy for a particular patient, depending on her personal characteristics and hospital data.

In many fields, the use of datasets contributed to develop effective solutions for problematic situations. There are different types of datasets that can be deployed. For example, online datasets, generally created by experts, public or private organizations and that are accessible (for free or upon payment) on their websites; machine-generated datasets to help users to solve specific problems or to retrieve specific data; and datasets related to a specific profession or organization, and that contain contextual data based on their history. Over the years, the latter type of datasets has been used in various fields to analyze actual data and obtain valuable knowledge to handle particular situations. In the medical and health-care field, clinical-administrative datasets would allow for capitalization, management and retrieval of relevant information [17].

The objective of this work is to create and analyze a real dataset on the clinical and personal characteristics (e.g. patient age, cancer stage at diagnosis, cancer subtype) of 601 patients with breast cancer and hospitalized, and to identify the different chemotherapeutic regimens used. The findings of this study would help to: (i) define the most appropriate breast cancer treatment depending on the individual characteristics of patients; (ii) address the specific needs of each patient; (iii) facilitate the design of personalized treatment for breast cancer patients; (iv) develop a predictive algorithm to forecast the stock of chemotherapeutic molecules; and (v) better use the limited resources. The key is to have a database that can be used to realize these ideas. To the best of our knowledge the created dataset and this study has not been proposed before.

II. Materials and Methods

A. Study Context and Sampling

In Morocco, the health care system consists of three different sectors: private non-profit, private for-profit (i.e. private clinics) and public sectors (i.e. government hospitals). Within the public sector, University Hospital Centers (UHCs) are the best-equipped facilities in terms of bed capacity and medical equipment. They include several medical disciplines and are involved in medical training and research. Thus, many Moroccan patients tend to go to UHCs for diagnosis and treatment. Based on the above, Mohammed VI UHC was chosen to perform the following investigation. The study was conducted from 2018, and included all the patients that were admitted to the Oncology and Hematology Center (OHC) of Mohammed VI UHC for breast cancer chemotherapy.

B. Data Collection

The data collection procedure is summarized in Fig. 1. Patients admitted to the OHC were categorized according to cancer type. Only breast cancer patients to whom chemotherapy was recommended were included in this study. Cancer stage was not an inclusion/exclusion criterium. However, patients who did not receive previous adjuvant or palliative chemotherapy were not included. Based on the European Society for Medical Oncology (ESMO) clinical practice guidelines for breast cancer treatment [18], the National Comprehensive Cancer Network (NCCN) guidelines [19], and at the medical oncologist's discretion, patients with the following criteria were not included:

- Early-stage luminal A breast cancer that has not spread to lymph nodes or affected 1-3 axillary nodes, with low clinical risk of recurrence, including tumor size and histological grade.
- Node negative pT1a, triple-negative breast cancers < 5mm in diameter.
- Node negative pT1b, HER2-positive or triple negative breast cancers.
- Patients with hormone receptors positive, HER2-negative and metastatic breast cancer who received endocrine therapy alone during the inclusion period.

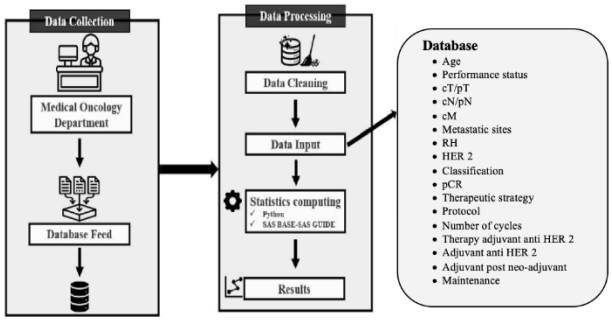

Fig. 1. Flow scheme of the study.

Data were recorded for patient demographics, breast cancer-related information including clinical and histopathological features, treatment options for patients and determinant factors that may affect or facilitate the decision-making process on whether or not to start a chemotherapy regimen, regardless if it is in the neoadjuvant, adjuvant or palliative setting. For each patient, therapy details were collected (i.e. schedule, number of cycles and the chemical dose administered). The performance status (PS) of each patient was evaluated and noted from 0 to 3 according to the Eastern Cooperative Oncology Group (ECOG) scale, with 0 being fully functional patient and 3 refers to bedridden patients. Tumor staging was assessed and classified according the American Joint Committee on Cancer (AJCC) classification system.

C. Data Analysis

Several languages including Python Program Language (PPL) were used to analyze, mine and synthetize the dataset. Statistical Analysis System (SAS) was used to generate correlation coefficients among variables and conduct a

principal component analysis (PCA) on the collected data. Prior to SAS analysis, all qualitative variables were converted into quantitative variables. Matplotlib and Seaborn were used for data visualization.

D. Ethical Considerations

Data were provided by the management of the OHC of Mohammed VI UHC following the approval of the head of the Medical Oncology Department. During our investigation, the identity of patients was never revealed.

III. Results

From January, a total of 739 breast cancer patients were admitted to the OHC of Mohammed VI UHC. This accounted for 25% of the total new cancer cases. Based on the inclusion criteria, 601 patients were selected for assessment. The mean age of the patients was 48.84 years (standard deviation, 11.12; median, 48.74 years; range, 24-85 years). At the time of diagnosis, 240 patients were premenopausal, which accounted for 39.93% of the total patients. Analysis of the data showed that most patients (57.40 %) were diagnosed in locally advanced stage or disseminated stages of disease and thus required systemic chemotherapy as the mainstay of treatment. All patients had a PS score of 0 or 1 on the ECOG scale. Patients with such scores are able to withstand chemotherapy. Indeed, clinical use of chemotherapy should be restricted to medically fit patients with the best PS scores and able to tolerate aggressive therapeutic regimens.

Based on our analysis, four main subtype profiles were found: hormone receptor positive and HER2 negative (54.47 %, n = 329), hormone receptor positive or negative and HER2 positive (25.29%, n = 152), and triple negative (TNBC) (19.80 %, n = 119). Pathologic complete response (pCR) was achieved in 23 patients (3.82%). All the patients included in the dataset received chemotherapy in either the adjuvant setting (67.22 %, n = 404), neo-adjuvant setting (18.63%, n= 112), adjuvant/post-neoadjuvant setting (1.99 %, n= 12) or in palliative intent (11.64 %, n = 70) (Fig. 2).

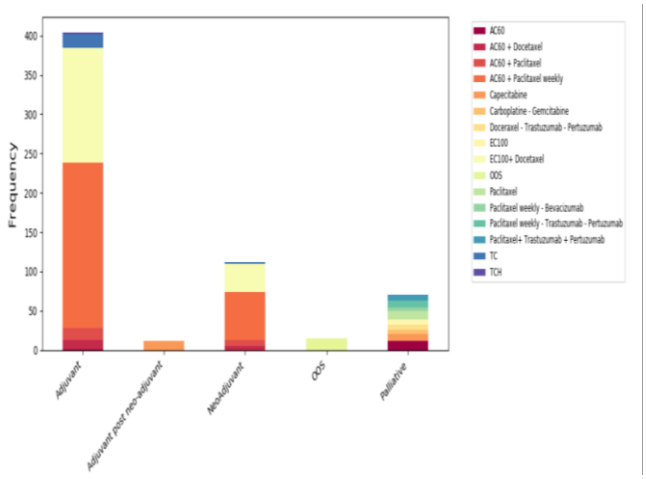

Fig. 2. Therapeutic strategy according to the prescribed treatments.

The distribution of patients according to their age was normal (Fig. 3).

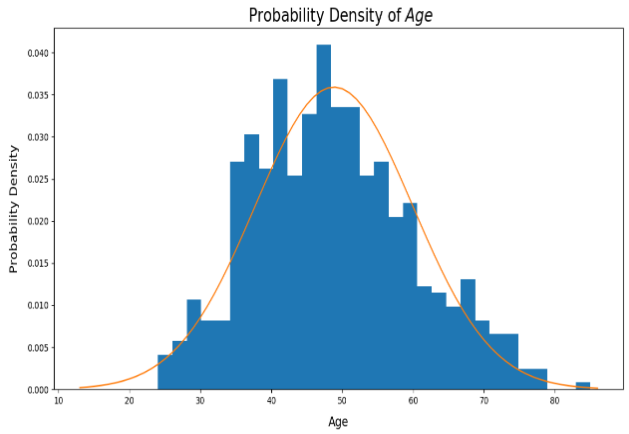

Fig. 3. Distribution of the variable age.

The distribution of patient age as a function of treatment regimen was also normal (Fig. 4). Sequential anthracycline and taxane-based regimen was the most prescribed therapy of curative intent (Fig. 4). Based on our findings, patient age appears to be a determinant factor for therapeutic decisions. Overall, the median number of pre- and postoperative chemotherapy cycles was 6, with a range of 4-8.

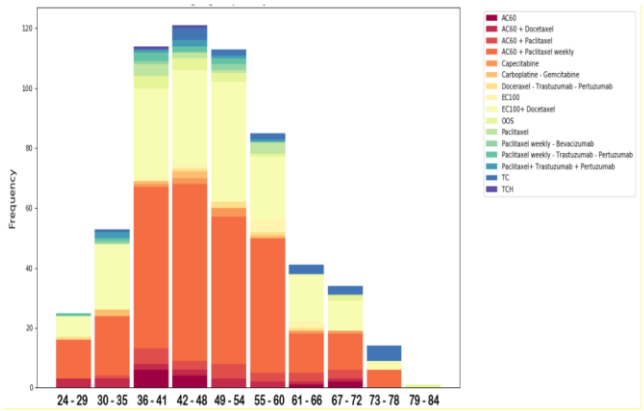

Fig. 4. Prescribed treatments according to patient age.

The subset of patients as defined by the AJCC was as follows: stage I (7.32 %, n = 44), stage IIA (18.30 %, n = 110), stage IIB (13.31 %, n = 80), stage IIIA (22.96 %, n = 138), stage IIIB (11.98 %, n = 72), stage IIIC (10.65 %, n = 64), and stage IV (11.81 %, n = 71). On the other hand, the cancer development stage of 22 patients (3.66%) could not be classified according to the AJCC criteria due to diagnostic failure or lack of information. As expected, the subtype classification of breast cancer was not correlated with anatomical prognostic factors, but can predict the tumor biological behavior.

The prescribed treatments for patients with first-line stage IV breast cancer were as follows: AC 60 was prescribed for 12 patients (17.14 %); Paclitaxel, Trastuzumab and Pertuzumab were prescribed for 11 patients (15.71 %); Paclitaxel, Pertuzumab and weekly Trastuzumab were prescribed for 9 patients (12.86 %); Paclitaxel and weekly-Bevacizumab were prescribed for 8 patients (11.43 %); EC100 was prescribed for

7 patients (10 %); Paclitaxel was prescribed for 7 patients (10%); Carboplatine and Gemcitabine were prescribed for 6 patients (8.57 %); Doceraxel, Trastuzumab and Pertuzumab were prescribed for 6 patients (8.57%); and Capecitabine was prescribed for 4 patients (5.71 %) (Fig. 2).

Fig. 5 shows the results of correlation among the different variables. According to practicing physicians, the variables presented in Fig. 5 are the most important ones for treatment prescription.

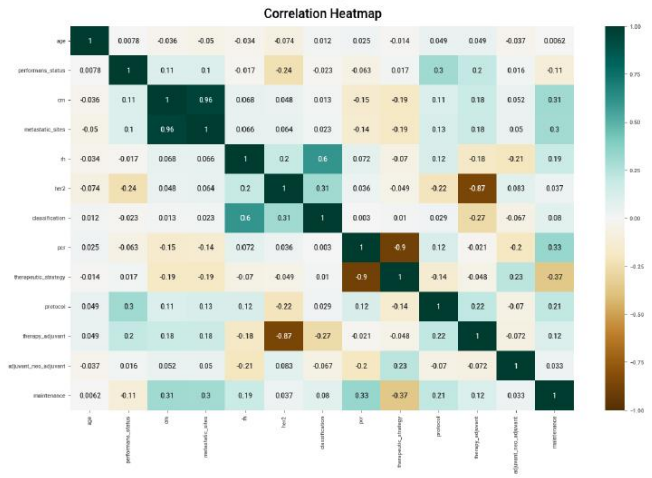

Fig. 5. Principal component analysis of causal attributions of breast cancer.

The correlation circle (Fig. 6) shows the distance between variables. Based on Fig. 5 and 6, it could be concluded that maintenance, therapeutic strategy, metastatic sites, cM and cT/pT were the main variables positively correlated with component 1 while HER2 and therapy adjuvant anti-HER2 were the main variables positively correlated with component 2. On the other hand, classification was the main variable negatively correlated with component 2. The findings of this study would be very useful in developing a machine learning model for the prediction of breast cancer treatment under Moroccan circumstances.

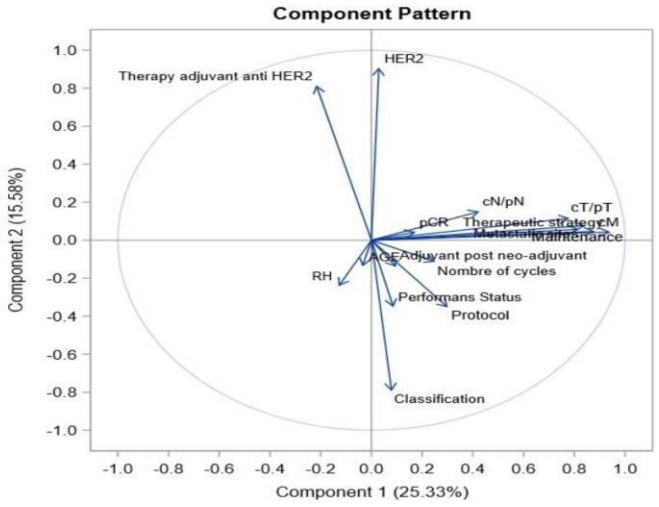

Fig. 6. Correlation circle of causal attributions of breast cancer.

IV. Discussion

Breast cancer is one of the most common cancers diagnosed in women. This disease causes a devastatingly high number of deaths every year [20]. In developing countries, controlling breast cancer is more problematic due to the lack of medical and human resources [21]. Furthermore, patients in developing countries have limited access to early diagnosis and to modern diagnostic tools [22]. In such circumstances, developing treatment prediction strategies could substantially contribute to save thousands of lives each year, especially when the tumor is detected in an advanced stage. The present study was conducted to analyze personal and medical data of breast cancer patients in Morocco. This would help to develop a predictive algorithm to determine the most appropriate treatment for patients, to better use the limited resources and to accurately estimate hospital needs.

In recent years, the use of datasets to assist in the decision making process has known a great success. This was observed in different fields including the medical one (e.g. [23], [24], [25]). Indeed, in the medical and health-care field, datasets are increasingly involved in planning and managing limited resources and to assist the decision-making process. According to [26], medical datasets are highly relevant for cancer research and clinical application. Medical datasets are composed of records collected during the clinical care process, and generally include individual records of patients, medical diagnoses, drug prescription and treatments [27]. In Morocco, the use of datasets to manage resources and take decisions is still in its infancy. Collecting, analyzing and processing breast cancer patient data would help in better utilizing the limited human and medical resources, and could improve the decision-making process depending on personal and medical records of patients.

The findings of this investigation showed that most breast cancer patients admitted to Mohamed VI UHC were 24-85-year-old, with a mean age of 48.84 years. Population-based data (271,173 patients) from Surveillance, Epidemiology, and End Results (SEER) cancer registries showed that more than 94% of breast cancer patients registered in 18 different USA cities/states and diagnosed between 2010 and 2015 were over 40 years old, with a mean age of 60 years [28]. Similarly, [25] found that among 235,368 French patients diagnosed with early breast cancer between 2011 and 2017 (identified by the Oncology Data Platform -(ODP) of the French National Cancer Institute (INCa)), 95% were 40 years old or above at the time of diagnosis. [29] used the Queensland Cancer Registry and found that among 3,079 women diagnosed with breast cancer in Queensland (Australia) between July 1st 2011 and June 30th 2012, 87.9% were older than 45 years. Based on our findings and the above-mentioned reports from the literature, it could be assumed that most women diagnosed with breast cancer are over 40 years of age. Up to date, the mechanisms underlying the relationship between age and breast cancer incidence remains unclear [30]. Indeed, the factors that induce breast cancer are complex, multiple and vary depending on the disease subtype [31][32]. According to many studies, age is not a determining factor that influences the incidence of breast cancer [33]. On the other hand, other studies indicated that in young patients, the pathophysiology of breast cancer is attributed to pathways linked to phosphoinositide 3-kinase,

immature mammary epithelial cells, growth factor signaling and mitogen activated protein kinase [34]. These diverging interpretations underline that research findings are not sufficiently conclusive regarding the likely effect of age on breast cancer incidence.

The findings of the present study showed that most women (45.59 %) had stage III breast cancer (based on the AJCC system), which is considered as locally advanced stage. Indeed, stage III is the stage where the tumor size is larger than 5 cm with involved lymph nodes [35]. In developing countries, most patients are diagnosed at advanced stages, which results in poor response to treatment. Examples from the literature showed that in developed countries most breast cancer cases are diagnosed at earlier stage. For example, [36] found that stage I tumors were the most common among 7,458 patients diagnosed at Asan Medical Center (Seloul, Korea) between January 1999 and December 2008. The researcher in [37] used data from 18 cancer registries (18 US cities/states) of the SEER program to determine the stage of breast cancer in 14,379 young women between 2010 and 2014. They found that most patients had tumors at stages I and II (28.2% and 45.2%, respectively), while only 19.0% had stage III breast cancer at diagnosis. The study [38] reported that among 1,703 breast cancer women diagnosed at the McGill University Health Centre (Montreal, Canada) between January 2010 and December 2013, the majority had stage I cancer. Stage I breast cancer is the stage where the tumor size is no more than 2 cm, with no involved axillary lymph nodes and without metastatic disease [39]. The fact that in Morocco most breast cancer patients are diagnosed at stage III raises the concern of late diagnosis, which would affect the efficiency of treatments and recovery rates. Taking this into account along with the limited resources in developing countries, developing efficient methods to determine the most effective treatment for each patient seems crucial.

Our findings showed that 11.81% of patients had a stage IV tumor. Stage IV cancer is that when the tumor has spread to other organs of the body (i.e. distant metastatic disease) [39]. Data from Mohamed VI UHC revealed that at this stage, nine different treatments were prescribed, with percentages varying from 5.71 to 17.14 %. This large number of treatments with close percentages of prescription may reflect the challenge of identifying the most effective treatment for stage IV breast cancer and highlights the importance of using datasets to decide on the most appropriate treatment depending on patient characteristics. Our findings also revealed that most patients have received systemic chemotherapy as the mainstay of treatment. Some currently well-established treatment methods that include escalating, de-escalating or omitting cytotoxic chemotherapy are not applicable in our context due to limited or no access to precision medicine. For example, the currently available gene expression signatures used to assess the risk of recurrence and the benefit of adjuvant chemotherapy in intermediate risk luminal breast cancers are not freely available in Morocco. Therefore, physicians rely solely on clinical-pathological factors, which lead to over-prescription of chemotherapy. Correlations among the variables included in the dataset confirm the fact that the treatment regimen depends mainly on cancer stage, which relies on clinical-pathological features.

Our study showed that the most prescribed treatment was the combination of Paclitaxel, Trastuzumab and Pertuzumab. Paclitaxel is a natural anticancer agent widely used to treat cancer patients due to its particular mechanism of action and successful outcomes [40]. Trastuzumab is a monoclonal antibody generally used to treat some HER2-positive tumors such as metastatic gastric cancer and breast cancer [41] [42][43]. This agent is known for its high cost, and in many countries it is not covered by health insurance. Consequently, in many countries patients cannot afford it [41]. Regarding Pertuzumab, it is an HER dimerization inhibitor and a major drug used to treat early and HER2-positive metastatic breast cancers [44][45]. The research [46] examined data from the Ontario Cancer Registry between April 1st 2012 and March 31st 2016 and found that, among 190 triple-negative breast cancer patients at stage IV, 25.3% have underwent surgery, 72.6% received systemic therapy and 58.9% received radiotherapy. They also indicated that the top drug regimens include anthracyclines and/or taxanes. On the other hand, these authors found that the cost of treatment for stage IV patients was four times higher than that of stage I-III patients. Based on the National Cancer Dataset (NCDB) data from 2010 to 2015, [47] compared the effects of different treatments on stage IV breast cancer patients from the US (i.e. 12,838 women who lived longer than six months after their diagnosis) and found that the most effective treatments were based on the combination of either systemic therapy (i.e. chemotherapy, endocrine therapy, or both) and surgery, or systemic therapy, surgery and radiation as compared to systemic therapy alone. These findings from the literature showed that the treatment for stage IV breast cancer may vary from systemic therapy, radiotherapy, surgery or their combination. In Morocco, while only conventional chemotherapeutic agents are either freely available or are covered by health insurance, the current challenges are to determine the most appropriate treatment for each patient and to avoid that demand outweighs supply. The use of datasets would considerably help in determining the most effective treatment depending on patient characteristics and tumor phenotype.

V. CONCLUSION

The findings of the present study provided valuable data on breast cancer patients in Morocco and the treatment regimens used. Similarly to many other developing countries, breast cancer is generally diagnosed at an advanced stage (i.e. stage III and stage IV) in Morocco. Systemic chemotherapy was found to be the mainstay of treatment. Nine different treatments were prescribed. The chemotherapeutic regimens used reflect the lack of a single standard or optimized treatments for patients. Further research could focus on evaluating the outcome of the different treatments, and to develop a predictive algorithm to determine the most appropriate treatment for each patient depending on demographic and medical characteristics. This would help to rationally use the limited medical and human resources and to improve patient outcomes and satisfaction.

VI. DECLARATIONS

A. Funding

This study did not receive any support from external funding agencies.

B. Conflict of Interest

The authors declare they have no conflict of interest.

C. Data Availability

Data available upon request from the corresponding author.

D. Ethics Approval and Consent to Participate

Human Research Ethics approval was obtained from the Mohamed VI UHC, Medical Oncology Department.

REFERENCES

[1] WHO (2021) World Health Organization. Breast Cancer. https://www.who.int/news-room/fact-sheets/detail/breast-cancer#:~:text=Scope%20of%20the%20problem,the%20world's%20most%20prevalent%20cancer. Accessed Juin 08th 2022.

[2] Susini T, Biglia N, Bounous VE (2022) Prognostic factors research in breast cancer patients: new paths. Cancers 14:971. https://doi.org/10.3390/cancers14040971.

[3] Park SK, Min YH, Lee M, Lee SB (2021) Causal attributions in breast cancer patients planning to undergo adjuvant endocrine therapy. Int J Environ Res Public Health 18:5931. https://doi.org/10.3390/ijerph18115931.

[4] Bessonneau V, Rudel RA (2020) Mapping the human exposome to uncover the causes of breast cancer. Int. J. Environ. Res. Public Health 17:189. https://doi.org/10.3390/ijerph17010189.

[5] Lee Y, Jeon Y-W, Im E-O, Baek J-M (2021) Causal attributions and quality of life of Korean breast cancer survivors. Asian Nurs Res 15:53–59. https://doi.org/10.1016/j.anr.2020.11.002.

[6] Seo J, Guk G, Park S-H, Jeong M-H, Jeong J-H, Yoon H-G, Choi K-C (2019) Tyrosine phosphorylation of HDAC3 by Src kinase mediates proliferation of HER2-positive breast cancer cells. J Cell Physiol 234:6428–6436. https://doi.org/10.1002/jcp.27378.

[7] Alphandéry E (2014) Perspectives of breast cancer thermotherapies. J Cancer 5:472–479. https://doi.org/10.7150/jca.8693.

[8] Khalis M, El Rhazi K, Charaka H, Chajès V, Rinaldi S, Nejjari C, Romieu I, Charbotel B (2016) Female breast cancer incidence and mortality in Morocco: comparison with other countries. Asian Pac J Cancer Prev 17:5211–5216. https://doi.org/10.22034/APJCP.2016.17.12.5211.

[9] Khalis M, El Rhazi K, Fort E, Chajès V, Charaka H, Huybrechts I, Moskal A, Biessy C, Romieu I, Abbass F, El Marnissi B, Mellas N, Nejjari C, Soliman AS, Charbotel B (2019) Occupation and risk of female breast cancer: a case-control study in Morocco. Am J Ind Med 62:838–846. https://doi.org/10.1002/ajim.23027.

[10] Charaka H, Khalis M, Elfakir S, Huybrechts I, Chami Khazraji Y, Lyoussi B, Soliman AS, Nejjari C (2021) Knowledge, perceptions, and satisfaction of Moroccan women towards a new breast cancer screening program in Morocco. J Canc Educ 36:657–663. https://doi.org/10.1007/s13187-019-01680-6.

[11] Yamada H, Kobayashi R, Shimizu S, Yamada Y, Ishida M, Shimoda H, Kato-Hayashi H, Fujii H, Iihara H, Tanaka H, Suzuki A (2020) Implementation of a standardised pharmacist check of medical orders prior to preparation of anticancer drugs to reduce drug wastage. Int J Clin Pract 74:e13464 https://doi.org/10.1111/ijcp.13464.

[12] Chen G, Xiao X, Zhao X, Tat T, Bick M, Chen J (2021) Electronic textiles for wearable point-of-care systems. Chem Rev 122:3259–3291. https://doi.org/10.1021/acs.chemrev.1c00502.

[13] Chen D, Si W, Shen J, Du C, Lou W, Bao C, Zheng H, Pan J, Zhong G, Xu L, Fu P, Fan W (2018) miR-27b-3p inhibits proliferation and potentially reverses multi-chemoresistance by targeting CBLB/GRB2 in breast cancer cells. Cell Death Dis 9:188. https://doi.org/10.1038/s41419-017-0211-4.

[14] Li J, Wang S, Wang Y, Wang X, Wang H, Feng J, Zhang O, Sun T, Ouyang O, Yin Y, Liu Y, Geng C, Yan M, Jiang Z (2017) Disparities of trastuzumab use in resource-limited or resource-abundant regions and its survival benefit on HER2 positive breast cancer: a real-world study from China. Oncologist 22:1333–1338. https://doi.org/10.1634/theoncologist.2017-0088.

[15] Chen Z, Ouyang O, Wang Y, Wang J, Wang H, Wu X, Zhang P, Huang J, Zheng Y, Cao W, Shao X, Xie N, Tian C, Liang H, Wang C, Zhang Y, Ren D, Wang X (2022) Real-world first-line treatment patterns and outcomes in hormone receptor-positive advanced breast cancer patients: a multicenter, retrospective study in China. Front Oncol 12:829693. https://doi.org/10.3389/fonc.2022.829693.

[16] Vokinger KN, Hwang TJ, Grischott T, Reichert S, Tibau A, Rosemann T, Kesselheim AS (2020) Prices and clinical benefit of cancer drugs in the USA and Europe: a cost–benefit analysis. Lancet Oncol 21:664–670. https://doi.org/10.1016/S1470-2045(20)30139-X.

[17] Owens M-R, Nguyen S, Karsy M (2022) Utility of administrative datasets and big data on understanding glioma treatment-a systematic review. Indian J Neurosurg. https://doi.org/10.1055/s-0042-1742333.

[18] Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, Zackrisson S, Senkus E (2019) Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 30:1194–1220. https://doi.org/10.1093/annonc/mdz173.

[19] NCCN (2021) Treatment by Cancer Type. NCCN Guidelines. https://www.nccn.org/guidelines/category_1. Accessed Juin 14, 2022.

[20] Raj S, Singh S, Kumar A, Sakar S, Pradhan C (2021) Feature selection and random forest classification for breast cancer disease. In: Satpathy R, Choudhury T, Satpathy S, Mohanty SN, Zhang X (eds) Data analytics in bioinformatics. Wiley: Hoboken, NJ, USA, pp 191–210. https://doi.org/10.1002/9781119785620.ch8.

[21] Zhuang Q, Wu J, Yu G (2021) A medical support system for prostate cancer based on ensemble method in developing countries. In: He X, Shao E, Tan G (eds) Network and parallel computing. NPC 2020. Lecture Notes in Computer Science, vol. 12639. Springer, Cham, pp 361–372. https://doi.org/10.1007/978-3-030-79478-1_31.

[22] Çakmak GK, Emiroğlu S, Sezer A, Canturk NZ, Yeniay L, Kuru B, Karanlık H, Soyder A, Gokgoz S, Sakman G, Ucuncu M, Akcay MN, Girgin S, Gurdal SO, Emiroglu M, Ozbas S, Bahadir A, Arici C, Toktas O, Demircan O, Çalık A, Polat AK, Maralcan G, Demirer D, Ozmen, V (2020) Surgical trends in breast cancer in Turkey: an increase in breast-conserving surgery. JCO Glob Oncol 6:285–292. https://doi.org/10.1200/JGO.19.00275.

[23] Mazo C, Kearns C, Mooney C, Gallagher WM (2020) Clinical decision support systems in breast cancer: A systematic review. Cancers 12:369. https://doi.org/10.3390/cancers12020369.

[24] Agossou C, Atchadé MN, Djibril AM, Kurisheva SV (2022) Mathematical modeling and machine learning for public health decision-making: the case of breast cancer in Benin. Math Biosci Eng 19:1697–1720. https://doi.org/10.3934/mbe.2022080.

[25] Dumas E, Laot L, Coussy F, Grandal Rejo B, Daoud E, Laas E, Kassara A, Majdling A, Kabirian R, Jochum F, Gougis P, Michel S, Houzard S, Le Bihan-Benjamin C, Bousquet P-J, Hotton J, Azencott C-A, Reyal F, Hamy A-S (2022) The French Early Breast Cancer Cohort (FRESH): a resource for breast cancer research and evaluations of oncology practices based on the French National Healthcare System Dataset (SNDS). Cancers 14:2671. https://doi.org/10.3390/cancers14112671.

[26] Lau EC, Mowat FS, Kelsh MA, Legg JC, Engel-Nitz NM, Watson HN, Collins HL, Nordyke RJ, Whyte JL (2011) Use of electronic medical records (EMR) for oncology outcomes research: Assessing the comparability of EMR information to patient registry and health claims data. Clin Epidemiol 3:259–272. https://doi.org/10.2147/CLEP.S23690.

[27] Hennessy S (2006) Use of health care datasets in pharmacoepidemiology. Basic Clin Pharmacol Toxicol 98:311–313. https://doi.org/10.1111/j.1742-7843.2006.pto_368.x.

[28] Kim HJ, Kim S, Freedman RA, Partridge AH (2022) The impact of young age at diagnosis (age <40 years) on prognosis varies by breast

cancer subtype: A U.S. SEER dataset analysis. Breast 61:77–83. https://doi.org/10.1016/j.breast.2021.12.006.

[29] Bates N, Callander E, Lindsay D, Watt K (2020) Patient co-payments for women diagnosed with breast cancer in Australia. Support Care Cancer 28:2217–2227. https://doi.org/10.1007/s00520-019-05037-z.

[30] Rozenblit M, Hofstatter E, Liu Z, O'Meara T, Storniolo AM, Dalela D, Singh V, Pusztai L, Levine M (2022) Evidence of accelerated epigenetic aging of breast tissues in patients with breast cancer is driven by CpGs associated with polycomb-related genes. Clin Epigenet 14:30. https://doi.org/10.1186/s13148-022-01249-z.

[31] Calaf GM, Ponce-Cusi R, Aguayo F, Muñoz JP, Bleak TC (2020) Endocrine disruptors from the environment affecting breast cancer. Oncol Lett 20:19–32. https://doi.org/10.3892/ol.2020.11566.

[32] Elaraby E, Malek AI, Abdullah HW, Elemam NM, Saber-Ayad M, Talaat IM (2021) Natural killer cell dysfunction in obese patients with breast cancer: A review of a triad and its implications. J Immunol Res 2021:9972927. https://doi.org/10.1155/2021/9972927.

[33] Wong FY, Tham WY, Nei WL, Lim C, Miao H (2018) Age exerts a continuous effect in the outcomes of Asian breast cancer patients treated with breast-conserving therapy. Cancer Commun 38:39. https://doi.org/10.1186/s40880-018-0310-3.

[34] Shin H-C, Han W, Moon H-G, Im S-A, Moon WK, Park I-A, Park SJ, Noh D-Y (2013) Breast-conserving surgery after tumor downstaging by neoadjuvant chemotherapy is oncologically safe for stage III breast cancer patients. Ann Surg Oncol 20:2582–2589. https://doi.org/10.1245/s10434-013-2909-6.

[35] Lee SB, Sohn G, Kim J, Chung IY, Lee JW, Kim HJ, Ko BS, Son BH, Ahn SH (2018) A retrospective prognostic evaluation analysis using the 8th edition of the American Joint Committee on Cancer staging system for breast cancer. Breast Cancer Res Treat 169:257–266. https://doi.org/10.1007/s10549-018-4682-5.

[36] Franzoi MA, Schwartsmann G, de Azevedo SJ, Geib G, Zaffaroni F, Liedke PER (2019) Differences in breast cancer stage at diagnosis by ethnicity, insurance status, and family income in young women in the USA. J Racial Ethn Health Disparities 6:909–916. https://doi.org/10.1007/s40615-019-00591-y.

[37] Savage P, Yu N, Dumitra S, Meterissian S (2019) The effect of the American Joint Committee on Cancer eighth edition on breast cancer staging and prognostication. Eur J Surg Oncol 45:1817–1820. https://doi.org/10.1016/j.ejso.2019.03.027.

[38] Russell CA (2003) Adjuvant systemic therapy for lymph node-negative breast cancer less than or equal to 1 cm. Curr Oncol Rep 5:72–77. https://doi.org/10.1007/s11912-003-0090-y.

[39] Tian X, Yu H, Li D, Jin G, Dai S, Gong P, Kong C, Wang X (2021) The mir-5694/Af9/Snail axis provides metastatic advantages and a therapeutic target in basal-like breast cancer. Mol Ther 29:1239–1257. https://doi.org/10.1016/j.ymthe.2020.11.022.

[40] Zhu L, Chen L (2019) Progress in research on paclitaxel and tumor immunotherapy. Cell Mol Biol Lett 24:40. https://doi.org/10.1186/s11658-019-0164-y.

[41] Sarosiek T, Morawski P (2018) Trastuzumab and its biosimilars. Polski Merkuriusz Lekarski 44:253–257.

[42] Akbari V, Chou CP, Abedi D (2020) New insights into affinity proteins for HER2-targeted therapy: beyond trastuzumab. Biochim Biophys Acta Rev Cancer 1874:188448. https://doi.org/10.1016/j.bbcan.2020.188448.

[43] Waller CF, Möbius J, Fuentes-Alburo A (2021) Intravenous and subcutaneous formulations of trastuzumab, and trastuzumab biosimilars: implications for clinical practice. Br J Cancer 124:1346–1352. https://doi.org/10.1038/s41416-020-01255-z.

[44] Robert M, Frenel JS, Bourbouloux E, Rigaud DB, Patsouris A, Augereau P, Gourmelon C, Campone M (2020) Pertuzumab for the treatment of breast cancer. Expert Rev Anticancer Ther 20:85–95. https://doi.org/10.1080/14737140.2019.1596805.

[45] Jagosky M, Tan AR (2021) Combination of pertuzumab and trastuzumab in the treatment of HER2-positive early breast cancer: A review of the emerging clinical data. Breast Cancer 13:393–407. https://doi.org/10.2147/BCTT.S176514.

[46] Brezden-Masley C, Fathers KE, Coombes ME, Pourmirza B, Xue C, Jerzak KJ (2020) A population-based comparison of treatment patterns, resource utilization, and costs by cancer stage for Ontario patients with triple-negative breast cancer. Cancer Med 9:7548–7557. https://doi.org/10.1002/cam4.3038.

[47] Stahl K, Wong W, Dodge D, Brooks A, McLaughlin C, Olecki E, Lewcun J, Newport K, Vasekar M, Shen C (2021) Benefits of surgical treatment of stage iv breast cancer for patients with known hormone receptor and HER2 status. Ann Surg Oncol 28:2646–2658. https://doi.org/10.1245/s10434-020-09244-5.