Exploring Breast Cancer Patterns for Different Outcomes using Artificial Intelligence

Nekane Larburu[1,2], Mónica Arrue[1,2] , Naiara Muro[1,2,3], Roberto Álvarez[1,2], Jon Kerexeta[1,2]
[1]eHealth and Biomedical Applications, Vicomtech, Donostia-San Sebastian, Spain
[2]Biodonostia, Donostia-San Sebastian, Spain
[3]Sorbonne Universités, UPMC Univ Paris 06, INSERM, Université Paris 13, Sorbonne, Paris Cité, UMR S 1142, LIMICS, Paris, France
nlarburu@vicomtech.org

***Abstract*—Breast Cancer is a complex disease characterized by multiple variables obtained from several data-sources, such as clinical, genetic or image sources. Over the past decades, various studies have tried to predict the outcome of breast cancer with the support of these data, and big advances have been done in this direction. However, only a few reports describe the causal relationships among the variables and outcomes, such as adverse events and survival rate, and usually they are very limited to a specific dataset. This research work presents a novel system that using data mining and visual analytics tools depicts in an intuitive way the patterns associated with different outcomes, such as treatment response and adverse events related to a treatment. For that the system processes heterogenous data coming from a real setting for primary breast cancer. This way clinicians can explore in a dynamic, fast and intuitive way whether certain group of patients are prone to certain outcomes.**

Keywords-breast cancer, pattern recognition, data mining, visual analytics

I. INTRODUCTION

Clinical guidelines are the narrative statements to assist clinicians about appropriate health care for a specific patient based on best available evidence [1]. However, often this evidence does not represent the real-world evidence, since not all type of patients can take part on the cohort studies, clinical trials etc., which are the foundation of clinical guidelines. Additionally, the fast progress in research studies, clinical guidelines often are not up to date, or may not consider some relevant information for certain group of patients [1].

Particularly, breast cancer is a complex disease that during the last few years has significantly improved its survival rates due to new discoveries in the genetic field and new drugs development. However, the real-world evidence is already implicit in the collected data from the real clinical practice. Therefore, the purpose of this research is to develop a dynamic and intuitive tool to discover new knowledge from clinical data of breast cancer patients. In particular, this paper focuses on the visualization of potential patters that might be associated to breast cancer outcomes, such as treatment response or adverse events related to a given treatment, and hence, give further insights to clinicians.

In this paper, we have used clinical data of primary breast cancer patients and implement data mining as well as visual analytics techniques within the European H2020 project DESIREE that aims to support Breast Units (BUs) in their decision-making process[1], not only with a clinical decision support system (CDSS), but also with visualization tools that can facilitate the extraction of new insights.

The reminder of this paper is organized as follows. In Section II, we present state of the art regarding data mining and visual analytics techniques for pattern recognition. Section III deals with the material and methods used to develop the solution. Section IV presents the obtained results in a specific use case, and Section V discussed the obtained results. Finally, Section VI concludes the paper with advantages and limitations of the developed tool, and future work.

II. STATE OF THE ART

Different techniques have been used over the years to aid clinicians in the estimation of the prognosis of a disease, and the appropriate visualization of these results.

Hence, this section presents A) Data Mining Techniques used in breast cancer and oncology to study different outcomes and B) Visual Analytics Tools existing in the literature with relevance for this research.

A. Data Mining Techniques

Current studies make use of several techniques to predict the outcome of a patient. These studies commonly relied on traditional statistical techniques such as regression models [2], but recently more sophisticated techniques based on machine learning, such as neural networks, have been also widely applied [3]–[8]. These studies usually present a model with own and specific datasets and their results are presented as the accuracy or sensitivity to predict outcomes, such as survival rates. As new studies also underline, the usage of other data sources, such as genetic data and image data can also improve the outcome prediction [9].

This study, instead of trying to predict the clinical outcomes, focuses on providing straightforward methods to extract new insights by depicting clinical data as patterns. Aligned with this approach, the study from Khalkhali et al [10] focuses on the hidden patterns extraction in breast

[1] http://www.desiree-project.eu/

cancer survival (yes/no) from an Iranian cohort study. Their study applies the classification and regression tree (CART) method in their cohort of 569 patients using 10-fold cross validation method. Their CART method achieves an accuracy, sensitivity and specificity of 80.3%, 93.5% and 53% respectively to predict whether the patient will survive or not. The study shows that the *Stage of Tumor* variable is the most important when predicting breast cancer survival.

Additionally, Zhen et al. [11] make a classifier to determine whether a tumor is benign or malignant, using the public Wisconsin Diagnostic Breast Cancer Dataset [12]. First, they extract the patterns of the benign and malign tumors separately, using K-means clustering method. They do not specify which the quarried patterns are. Once they have the patterns of each type of tumor, they apply Support Vector Machine (SVM) classifier. Their classifier obtains an accuracy of 97.38% using cross validation technique to avoid the overfitting. Their results demonstrate that the usage of patterns before applying SVM classifier improves the results.

Our study aims to determine over time, considering heterogenous data, which are the attributes that influence most different outcomes, and represent them in an intuitive and dynamic manner. Hence, although outcomes are needed to estimate and train our model, we do not aim to directly predict the survival probability or outcomes, but to detect whether there is a pattern associated to different outcomes.

B. Visual Analytics Tools

Due to the limited time of clinicians to explore data, it is essential to give them intuitive and fast tools to interpret them. In this context, visual analytics is the science of displaying information through interactive interfaces focused on analytical reasoning [13]. Analytical reasoning is the ability to detect patterns within the data and to gain deep insights by looking at the representation of large amounts of data. One of the most widespread criteria for classifying the visualization types is the dimensionality of the visualization, that is, the number of attributes that allows to show. Univariate visualization -one dimension- is the simplest form of data analysis and its goal is to gain insight about the distribution, the central tendency and the spread of an attribute. On the other hand, the main objective of the multivariate visualization -two or more dimensions- is to allow the analysis of the relationship or interaction between attributes [14], [15].

Multivariate analysis not only allows us to check the distribution of each of the variables, but also to analyze the relationships, patterns and correlations between these attributes. One of the best-known charts for visualizing multivariate data is the parallel coordinate plot [16], [17].

In this graph, each of the dataset attributes is represented by a vertical axis. These axes are positioned in parallel saving the same distance from each other. Each of the dataset samples is represented by a horizontal line so that it crosses the axes of the attributes taking the corresponding value. These lines are usually colored according to a criterion defined by the research question searched in these patterns. This type of graph is perfect for analyzing the relationship between many variables at once and find patterns in the data [18]. As presented in [17], some of the advantages are the integrity of results (i.e. different combination of results can be presented), connectivity of results (i.e. the relationship between results can be investigated easily), and consequences of results (i.e. provides new facilities to interact with data and augments the process of knowledge acquisition). But also has disadvantages, such as "over-plotting", which is equal to an occlusion problem, which occurs mainly when the dataset is medium-sized. The result of it is an image which is far too cluttered to be able to analyze any trends or structure, due to noisy intersections.

An example of this visualization technique is Fastbreak, a tool designed to facilitate the analysis and subsequent visualization of huge amounts of genomic data [19]. Specifically, it allows the analysis of structural variations of next-generation sequencing data. One of the visualizations that this tool provides is a parallel coordinates plot (Figure 1), where each of the vertical axes represents the level of expression that a gene may have in each particular cancer [20]. On the other hand, the horizontal lines represent a different gene and they are colored according to their genetic expression -red for the most expressed genes and blue for the least expressed-, which allows us to explore gene expression similarities between different cancers (Figure 1).

Epidemiology is another healthcare field that is also beginning to benefit from visual analytics, as it deals with big amounts of information related to the incidence, distribution, and other factors of a concrete disease in different populations. The parallel coordinates graph created by the German Centre for Cancer Registry Data called Flow of Cancer Statistics [21], [22] is an example of this. These parallel coordinates plot displays four different attributes: age rank, gender, tumor localization and implication (Figure 2). The tool allows to select a range on each parallel coordinate's axis to highlight a specific population from all the dataset. This feature facilitates the discovery of patterns between data and defining a specific patient profile.

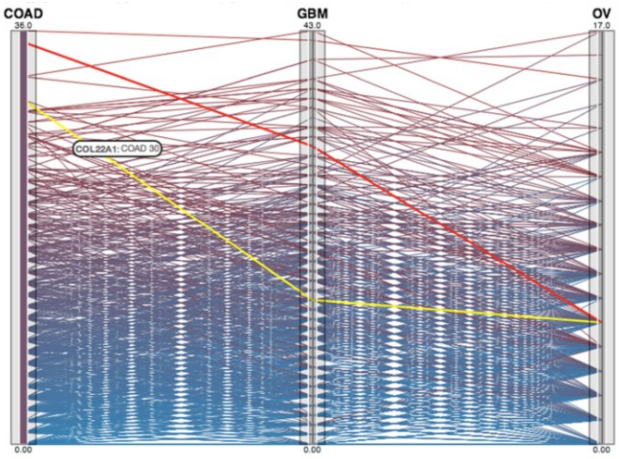

Figure 1. Two highlighted genes within the Fastbreak parallel coordinates plot displaying structural variations across colon adenocarcinoma (COAD), glioblastoma multiforme (GBM) and ovarian cancer (OV).

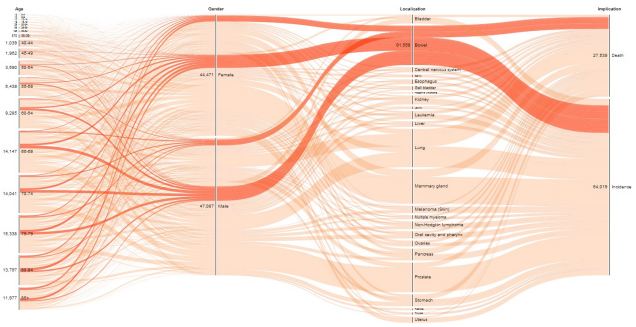

Figure 2. Flow of Cancer Statistics parallel coordinates plot where patients with bowel tumor have been highlighted.

Despite the potential limitations that the parallel coordinates visualization tool may have for low number of cases, the multidisciplinary team of clinical partners involved in DESIREE has selected it to be implemented in DESIREE as it one of the most intuitive tools for this type of pattern representation.

III. Material and Methods

This section presents first the clinical dataset used in this research study followed by the data mining methods applied for extracting patterns from it.

A. Dataset

Breast cancer is a complex disease that could be studied in different phases based on when the case is being discussed by the Breast Unit. In DESIREE the following phases or scenarios have been defined:

A. Diagnosis
B. After Neo Adjuvant Treatment (before surgery)
C. After Surgery (neo-adjuvant treatment)
D. After Surgery (no neo-adjuvant treatment)
E. Follow up

Depending on the scenario in which each patient is, different clinical data will be studied and different treatments supervised.

- Attributes to be studied.

The dataset used for this study was composed by real data of Primary Breast Cancer patients coming from four different European hospitals with an amount of 471 patients. It is composed by a subset of 148 attributes selected by a multidisciplinary team of clinicians and reviewed periodically. Most of these attributes are already stored in the Electronic Health Records (EHRs), but have been extended to include additional information that is of interest for clinicians. The types of attributes found in this dataset are:

- Categorical variables (e.g. tumor stage)
- Numerical variables: integer and float (e.g. tumor size)
- Logical variables (e.g. if it is a recurrent disease)

TABLE I presents the analysis of the values of the demographic data available in the dataset for the patients.

TABLE I: Demographic Data Description

Attribute	Mean ± SD/percentage
Age	58.21±13.18 (years)
Age at Menarche	12.9±1.5 years
Number of Abortions	0.43±1.24
Number of Pregnancies	2.23±2
Number of Live Birth	1.64 ± 1.19
Age at First Pregnancy	27.47 ± 7.44
Menopause status	Perimenopause: 5.9% Postmenopause: 58.6% Premenopause: 30% NA: 5.4%

*NA: missing values

Patients were treated with several therapies grouped in (i) surgical procedures, (ii) radiotherapy protocols, (iii) endocrine therapies and (iv) chemotherapy protocols, depending on the scenario and the status of the patient.

- Outcomes to be studied:

As explained in Section II. B, a pattern is analyzed according to a criterion. This criterion is what we consider to obtain the pattern. This study focuses on pattern discovery for the following criteria, called outcomes:

a) Guideline Compliance: since DESIREE implements a guideline-based CDSS, when BUs are not compliant with the guideline recommendations, this information is stored in the system, and is taken into account also to determine whether there is a tendency not to follow the guideline for some specific group of patients.

b) Adverse Events (AE) or Toxicities: is defined as "any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use of a medical treatment or procedure that may or may not be considered related to the medical treatment or procedure" [23]. One of the most relevant systems to determine the toxicities is the one developed by the US National Cancer Institute (NCI) and known as the NCI Common Terminology Criteria for Adverse Events or CTCAE2, and used in DESIREE.

c) Treatment Response: only neoadjuvant treatment can be assessed "objectively", and therefore, one of the objectives is to explore whether certain patients may have better or worse response to neoadjuvant treatments. The possible values that can take are (from worst to best): disease progression, stable disease, partial response and complete response.

d) Clinical Outcomes: the clinical outcome is measured by relapse and exitus (due to breast cancer or not) or survival (from which we can obtain the survival rate based on the diagnosed date and current date).

These are the most relevant outcomes that the multidisciplinary team of clinicians has selected to be studied. However, other aspects such as Patient Reported Outcomes (PRO), should also be considered [24] as they include a personal perception of the treatment effectivity from the patient point of view. But in the available use case this data is not retrieved, and therefore, it is not included.

B. Data Mining

This section focuses on explaining the workflow performed to calculate the most relevant attributes so as to get a clear visualization of those patterns that could better give insights about a specific outcome.

1) Preprocessing

Data understanding and data preparation stages are among the most important steps in the data mining applications.

Firstly, the category type of the attributes is detected, to treat each group differently depending on if they are categorical, numerical or boolean. Secondly, some variables have been excluded if they are not representative (for the preliminary results). For example, if a categoric attribute takes the same value for all patients, this attribute will not provide additional information, and hence, it is not studied (e.g. if all the patients have the cancer in the 'left' breast, 'breastSide' variable is not considered as it does not give any additional information that will affect the results).

Additionally, the problem caused by missing values is also studied. The imputation of the missing values could cause "noise" within the dataset. Hence, only the non-null values are used when applying the statistical methods to estimate the correlation between two variables, as explained in the following chapter, and null values are kept to "0".

Finally, all those attributes that have the same clinical meaning have been grouped together. Several attributes may come from different data sources but have the same clinical meaning (e.g. tumor size attribute may come from the ultrasound, MRI and mammography). Therefore, all these clinical attributes have been grouped together into a single attribute to avoid data replication and missing values.

2) Statistical Analysis

Once the data is preprocessed, the main objective is to obtain the dependencies between all parameters and the studied outcome. To illustrate the obtained pattern in an intuitive way, all dependencies within all parameters are calculated using a *correlation matrix*. Besides capturing the most relevant parameters for a specific outcome, the dependencies between these relevant parameters for their later visualization purposes are also obtained (see Section IV. B).

To study patterns related to different outcomes of a patient, first the dataset is filtered based on the selected scenario. Additionally, clinicians can study the patterns considering all treatments together or by each treatment group: surgery, endocrine therapy, chemotherapy, and radiotherapy. Hence, different correlation matrixes are calculated for each possible treatment and scenario.

The correlations are updated frequently to cope with the new patients' data added to the system. Once the matrixes are created/updated, they are stored into a SQL database. To build the correlation matrix, the R package 'polycor' [25] is used. This package computes a heterogenous correlation matrix, consisting of Pearson product-moment correlations between numeric variables, polyserial correlations between numeric and categorical variables, and polychoric correlations between categorical variables, and returns a correlation level between -1 and 1. This way, it is possible to compare whether two variables are more correlated than others even when the type of variables are different.

When the number of instances shared between two attributes are not sufficient (e.g. there are less than 10% values in common that are not null) to determine the correlation between them, the correlation value of these is set to 0. This way the correlation is set as insignificant.

3) Main Attributes Selection

For selecting the main attributes, the scenario and treatment(s) to be analyzed must be defined by the user. Once we have the corresponding correlation matrix, the number of the most correlated attributes ("n") that match with the selected outcome are given.

Finally, to represent these attributes in an intuitive and attractive way, the main "n" attributes (i.e. the axes) are organized depending on their correlation level between them. That is, the "n" selected attributes are arranged in such way that the most correlated attributes are next to each other. This way the obtained pattern is illustrated more intuitively to the user.

IV. Preliminary Results in a Use Case

This preliminary study, although contains real clinical data, lacks from sufficient outcomes, which has a direct influence on the current results. However, the tool is finalized and when the number of available data increases new clinical knowledge about breast cancer will be extracted, so that clinicians can study potential hypothesis. As proof of concept this section presents the results obtained for scenario B - After Neo Adjuvant Treatment (before surgery) - and the criterion of the neoadjuvant treatment response. This criterion has categorical type and contains as possible values: (1) *Complete Response* when the disappearance of all target lesions occur, (2) *Partial Response* (PR) when at least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters, (3) *Progression* or Progressive Disease (PD), when at least 20% of increment in the sum of diameters of target lesions or absolute increment of at least 5 mm occur, and (4) *Stable Disease* when neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum diameters. In the visualization subsection, the corresponding parallel coordinates plot is shown.

A. Data Mining

For patients in scenario B, the main n = 3 attributes and their correlations related to neoadjuvant treatment response (categorical), are the following (Table2, Figure 3):

TABLE II: MAIN ATTRIBUTES CORRELATION

Attribute	Type of variable	Polyserial correlation
Number of live birth	Numerical (integer)	-0.25
Age at first pregnancy	Numerical (integer)	0.28
Tumor size	Numerical (float)	0.62

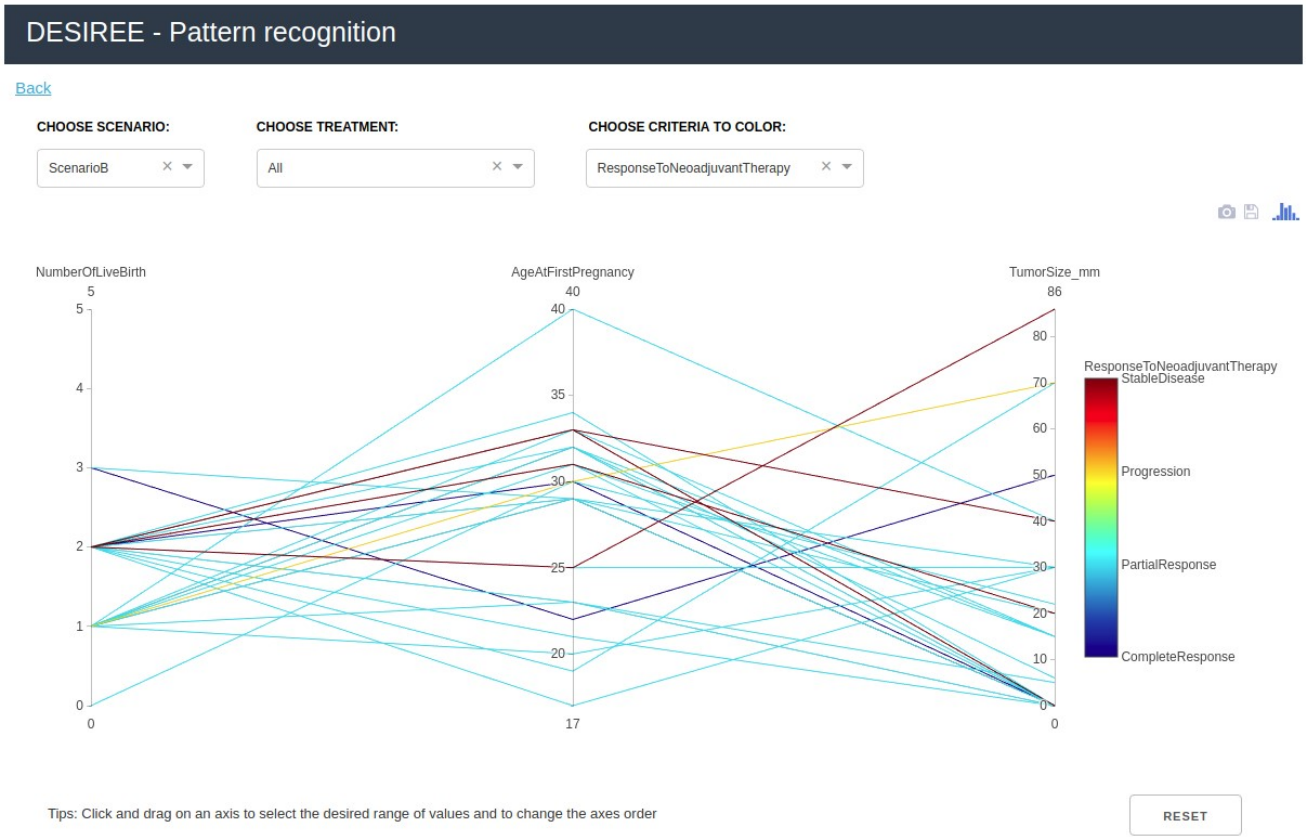

Figure 3. Parallel coordinates plot obtained for scenario B and the response to treatment criterion

B. *Visualization*

Figure 3 represents the most correlated attributes with the selected criterion (i.e. treatment response to neoadjuvant treatment) obtained in the data mining process. Each of the axes represents one of these attributes and each of the samples lines in horizontal represents a patient. These lines are coloured according to different values of the selected criterion. In red *Stable Disease*, in yellow *Progression*, in light blue *Partial Response* and in dark blue *Complete Response*.

Notice that only the samples (i.e. patients) that have the data fulfilled for the three obtained most correlated attributes (i.e. axes) and the selected criterion (i.e. color) are plotted. In the case that a sample does not have one of the required data, the line will not be drawn. Hence, in this case, for the scenario B and this treatment response criterion, just 68 patients (i.e. samples lines) are obtained.

Notice that the tool also gives clinicians the possibility of adding or removing attributes (i.e. axes) to the parallel coordinates plot. This way, although by default the most correlated attributes calculated by the data mining process are presented, clinicians can also explore further information.

V. Discussion

The purpose of this study is to develop a dynamic and intuitive tool for clinicians that allows the rapid inference of data to explore potential patters related to outcomes that can support their hypothesis. In the obtained results, although the number of lines is not very high, it is already possible to visualize some patterns and get some insights from the captured clinical data (Figure 3). However, it is expected to obtain more significant patterns with higher number of samples.

On the one hand, the tumor size is the attribute with higher correlation rate with neoadjuvant treatment response, which clinically seems relevant. Tumor size axis shows the range in which the data are found; this attribute ranges from zero millimeters to 85 millimeters. Looking at this axis, it can also be seen that most patients have a tumor size from zero to 30 millimeters. On the other hand, it can also be seen that most of the patients for whom there is data have a partial response to neoadjuvant treatment, since most of the lines are colored light blue.

VI. Conclusion

This is a novel approach in the field of breast cancer since advanced data mining methods are usually used to develop models that possess a high degree of predictive accuracy. But they are rigid and not accessible for clinicians

to train them dynamically to obtain new insights over time. Besides, to the best of our knowledge, there are no tools for clinicians to explore the results in a dynamic, intuitive and fast manner. Therefore, current studies carried by clinicians usually are high time consuming.

The presented tool allows them to explore easily all relevant information and check visually whether some pattern related to specific criteria exists or whether their hypothesis is reflected in the clinical data. However, one of the main pitfall of this tool is its limitation when some of the attributes to be illustrated contains null values. As mentioned above, only the samples that have the data fulfilled for the "n" obtained most correlated attributes (i.e. axes) – or the ones selected by clinicians – and the selected criterion (i.e. color) are plotted. Hence, to be able to visualize the results and gain insight on patterns and relationship between attributes, it is necessary to have sufficient number of samples "complete".

The current study requires the input of all values in DESIREE system manually, since it is not integrated with the four clinical partners involved in the project. This is a widely spread limitation since most EHR are not interoperable and the available data in current EHRs is not structured in a computer interpretable way and, consequently, it is not directly exploitable. In future work the system will be integrated with the EHR from each hospital. This will improve the availability of the data and outcomes, and consequently, the obtained results are expected to be more significant. Moreover, notice that even this tool gives clinicians the opportunity to explore potential relevant parameters that may have an impact on one of the outcomes, the statistically significance is not provided. Thus, future work will also consider this aspect.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 690238.

Additionally, the authors acknowledge support from the four clinical partners involved in the project: Onkologikoa, Eresa Grupo Médico, Hôpital Européen Georges-Pompidou and Hôpital Saint-Louis.

REFERENCES

[1] B. Séroussi, J. Bouaud, and É.-C. Antoine, "OncoDoc: a successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer," Artif. Intell. Med., vol. 22, no. 1, pp. 43–64, Apr. 2001.

[2] D. R. Cox, "Regression Models and Life-Tables," in Breakthroughs in Statistics, Springer, New York, NY, 1992, pp. 527–541.

[3] J. Kerexeta, A. Artetxe, V. Escolar, A. Lozano, and N. Larburu, "Predicting 30-day Readmission in Heart Failure using Machine Learning Techniques:," 2018, pp. 308–315.

[4] H. B. Burke et al., "Artificial neural networks improve the accuracy of cancer survival prediction," Cancer, vol. 79, no. 4, pp. 857–862, 1997.

[5] E. Biganzoli, P. Boracchi, L. Mariani, and E. Marubini, "Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach," Stat. Med., vol. 17, no. 10, pp. 1169–1186, 1998.

[6] D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods," Artif. Intell. Med., vol. 34, no. 2, pp. 113–127, Jun. 2005.

[7] A. Al-allak, L. Intabli, G. Bertelli, and P. Lewis, "Artificial intelligence: A new generation of intelligent predictive models to guide adjuvant treatment decisions for patients with breast cancer?," Eur. J. Surg. Oncol., vol. 44, p. S43, 2018.

[8] K. Shaffer, Can Machine Learning Be Used to Generate a Model to Improve Management of High-Risk Breast Lesions? Radiological Society of North America, 2018.

[9] D. Sun, A. Li, B. Tang, and M. Wang, "Integrating genomic data and pathological images to effectively predict breast cancer clinical outcome," Comput. Methods Programs Biomed., 2018.

[10] H. R. Khalkhali, H. Lotfnezhad Afshar, O. Esnaashari, and N. Jabbari, "Applying data mining techniques to extract hidden patterns about breast cancer survival in an Iranian cohort study," J. Res. Health Sci., vol. 16, no. 1, pp. 31–35, 2016.

[11] B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms - ScienceDirect," vol. 41, no. 4, pp. 1476–1482, 2013.

[12] "UCI Machine Learning Repository: Breast Cancer Wisconsin (Original) Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original). [Accessed: 20-Jun-2018].

[13] R. May, P. Hanrahan, D. A. Keim, B. Shneiderman, and S. Card, "The state of visual analytics: Views on what visual analytics is and where it is going," in 2010 IEEE Symposium on Visual Analytics Science and Technology, 2010, pp. 257–259.

[14] K. Nazemi, Adaptive Semantics Visualization. Springer, 2016.

[15] J. Brownlee, "Better Understand Your Data in R Using Visualization (10 recipes you can use today)," Machine Learning Mastery, 29-Jan-2016. .

[16] D. (DJ) Sarkar, "The Art of Effective Visualization of Multi-dimensional Data," Towards Data Science, 15-Jan-2018. [Online]. Available: https://towardsdatascience.com/the-art-of-effective-visualization-of-multi-dimensional-data-6c7202990c57. [Accessed: 08-Jun-2018].

[17] A. Cuzzocrea and D. Zall, "Parallel Coordinates Technique in Visual Data Mining: Advantages, Disadvantages and Combinations," in 2013 17th International Conference on Information Visualisation, 2013, pp. 278–284.

[18] "Parallel Coordinates Plot - Learn about this chart and tools." [Online]. Available: https://datavizcatalogue.com/methods/parallel_coordinates.html. [Accessed: 07-Jun-2018].

[19] "Fastbreak."[Online]. Available: http://fastbreak.systemsbiology.net/. [Accessed: 14-Jun-2018].

[20] R. Bressler et al., "Fastbreak: a tool for analysis and visualization of structural variations in genomic data," EURASIP J. Bioinforma. Syst. Biol., vol. 2012, p. 15, Oct. 2012.

[21] "Visualizing Cancer Data through Flows." [Online]. Available: http://www.visual-telling.com/vis/cancer_statistics/index.html. [Accessed: 07-Jun-2018].

[22] O. Bieh-Zimmert, "Flow of Cancer Statistics | Visual Telling." .

[23] C. L. Shapiro and A. Recht, "Side effects of adjuvant treatment of breast cancer," N. Engl. J. Med., vol. 344, no. 26, pp. 1997–2008, Jun. 2001.

[24] N. Muro, N. Larburu, J. Bouaud, and B. Seroussi, "Weighting Experience-Based Decision Support on the Basis of Clinical Outcomes' Assessment.," Stud. Health Technol. Inform., vol. 244, pp. 33–37, 2017.

[25] J. Fox, "polycor: Polychoric and Polyserial Correlations." 2016.