Intelligent Clinical Decision Support Systems for Patient-Centered Healthcare in Breast Cancer Oncology

Boomadevi Sekar[a], Jean-Baptiste Lamy[b], Naiara Muro[b, c, e], Amaia Ugarriza Pinedo[d], Brigitte Seroussi[b]

[a]School of Computing, Ulster University, Belfast, UK

[b]Sorbonne Universités, UPMC Univ Paris 06, INSERM, Université Paris 13, Sorbonne
Paris Cité, UMR S 1142, LIMICS, Paris, France
b.sekar@ulster.ac.uk, jean-baptiste.lamy@univ-paris13.fr, nmuro@vicomtech.org,
augarriza@bilbomatica.es, brigitte.seroussi@aphp.fr

Nekane Larburu[c, e], Gilles Guézennec[b], Jacques Bouaud[b], Frank Guijarro Masero[d], Mónica Arrúe[c, e], Hui Wang[a]

[c]Vicomtech, Donostia-San Sebastian, Spain
[d]Bilbomática, Bilbao, Spain
[e]Biodonostia, Donostia-San Sebastian, Spain
nlarburu@vicomtech.org, gilles.guezennec@univ-paris13.fr, jacques.bouaud@aphp.fr,
frank@bilbomatica.es, marrue@vicomtech.org,
h.wang@ulster.ac.uk

Abstract— **Breast cancer is the most common type of cancer in women worldwide, with incidence rate being second highest to other types of cancer. In the current clinical setting, multidisciplinary breast units are introduced to improve the quality of the therapeutic decision based on the best evidence-based practices. DESIREE project aims to provide a web-based software ecosystem for personalized, collaborative and multidisciplinary management of primary breast cancer by specialized breast units, from diagnosis to therapy and follow-ups. In order to provide a multi-model decision support to clinicians in present clinical settings, the project develops and integrates three modalities of decision support, namely guideline-based decision support system (DSS), experience-based DSS and case-based DSS. Visual analytics GUI are developed to properly adapt the results of the DSSs and graphically represent them to the clinician in a user-friendly manner. DESIREE information management system (DESIMS), serves as the interface between the user and DSSs for entering patient data and viewing the results in the visual analytics GUI. In this paper, we present the overall architecture, workflow and integration of the three DSSs in the DESIREE platform.**

Key words: Clinical decision support system; ontology; clinical based guidelines; experience-based decision support; case-based decision support

I. INTRODUCTION

Breast cancer is the most common type of cancer among women, with 1.6 million women diagnosed every year [1]. Although the increase in life expectancy could contribute to the increased incidence rate, according to Cancer Research UK, when compared to the statistics from previous 15 years, there is an 11% increase in the cases diagnosed under the age of 50 [2]. However, despite the increased incidence rate, women dying from breast cancer aged under 50 has reduced, which is a reflection of increased awareness, increased rates of diagnosis, advancement in screening techniques, and improvement in drug treatments [2]. In current practice, multidisciplinary breast units (MBUs) guided by clinical practice guidelines (CPGs) are introduced to improve the quality of the therapeutic decision with adherence to the best evidence-based practices. However, two problems need to be considered in such evidence-based practice. First, the MBUs are required to handle vast amount of patients and clinical data, and thereby may be often limited on the amount of time spent to review each individual patient. Second, CPGs work for "standard" patients, but provide inadequate solutions to outliers (complex cases).

To address the first problem, research shows [3] that implementation of CPG based clinical decision support systems (CDSSs) providing evidence-based recommendations has gained various benefits, including reducing the risk of medical errors, improving efficiency and patient throughput, and facilitating the process of every day patient care practice. Literature shows that evolution of research in the development of CDSS has come a long way, from using Bayesian probability in 1960s, to knowledge encoded into if-then rules in MYCIN in 1975, to using Bayesian networks in late 1980s and algorithms for learning Bayesian networks from data in 1990s, to using statistical and machine learning methods in the past few decades. A detailed survey on the state-of-art CDSS developed is presented in [4] and [5], where it is emphasized that yet great challenges exist in building reliable, efficient, and user-friendly CDSSs.

Now, to address the second problem, literature shows [6] that the current research is directed towards the development of patient-centered healthcare with the aid of big data analytics, which aims to customize treatment for an individual patient based on their likelihood of response to a therapy. Long-term goals of such research direction are numerous, including disease delineation and stratification, early detection of symptoms and diseases, and even identification of diseases in an individual well before the symptoms actually occur. Thus, we believe implementing intelligent CDSSs for patient-centered healthcare at the MBUs, could not only lower the burden on the clinicians and benefit from the advantages of evidence-based medicine, but can also allow personalised healthcare decision making possible.

DESIREE (Decision Support and Information Management System for Breast Cancer), a European Union funded project,[1] is aimed at such direction in developing a web-based software ecosystem for the personalized, collaborative, and multidisciplinary management of primary breast cancer (PBC) by MBUs. DESIREE is aimed at developing an intelligent CDSS that would not only provide decision support based on the available therapy options from the existing CPGs, but will also incorporate experience-based and case-based decision support modalities that are built based on evolving knowledge acquired from previous patient cases. In this paper, we will present the overall architecture and workflow of the proposed CDSS framework for DESIREE.

II. CLINICAL DECISION SUPPORT SYSTEM FOR DESIREE

One of the main objectives of the DESIREE project is to provide an efficient and reliable decision support for the therapeutic decisions made by the MBUs, including surgery, radiotherapy and systemic therapies. With the aim to provide a personalized state-of-the-art clinical decision support to the MBUs, the project develops and integrates three modalities of decision support, namely guideline-based DSS (GL-DSS), experience-based DSS (EX-DSS) and case-based DSS (CB-DSS).

Fig. 1 shows the high-level framework of the three modalities of decision support in the DESIREE platform. The framework contains the DESIREE information management system (DESIMS), fast healthcare interoperability resources (HAPI-FHIR) server, the three DSSs and the corresponding visual analytics GUI.
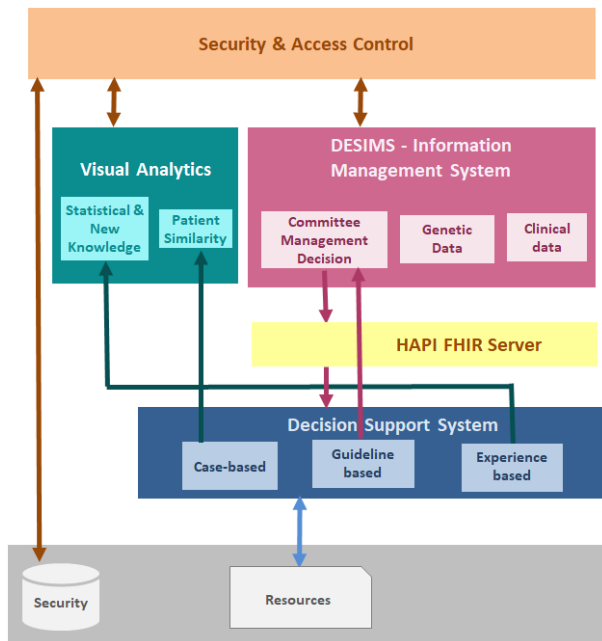

Figure 1. High-level framework DESIREE platform

A. DESIMS – Information Management System

DESIMS is the information management system, which receives and manages all the clinical data from MBUs. It contains a clinical interface through which clinicians from different medical specialties can enter patient data and store them according to a common data model. The clinical data model in DESIMS contains patient data from the MBUs. This includes demographic data, patient's physiological data, family history, image test results, neo-adjuvant, oncological, surgical and radiotherapy treatments, follow-ups and patient reported outcomes for most of the patients. The genetic data model on the other hand, contains genetic data for some selected patients.

To facilitate a secure access to the DESIREE platform, a security and access control system with authentication tokens and authorization checks is implemented. The DESIMS system offers a 'committee management decision', which gives access to the heterogeneous patient data and retrieves the results and therapeutic propositions from the three DSSs.

To manage the large volume of data and allow clinicians from different BUs to access them in a systematic way, and retrieve the relevant decision support from the DSSs, the clinical pathway of breast cancer patients is categorized into five scenarios, as listed below.

- Scenario A: After diagnosis
- Scenario B: After neo-adjuvant Therapy – No surgery
- Scenario C: After surgery – After neo-adjuvant Therapy
- Scenario D: After surgery – No neo-adjuvant Therapy
- Scenario E: After surgery and adjuvant treatment

B. HAPI-FHIR Server

The clinical data from DESIMS are transferred to the DSSs through HAPI-FHIR server, which is an open source standard framework developed for exchanging healthcare data in a systematic manner.

The database in FHIR server stores the clinical and genetic data of the patient in his/her current scenario, to be retrieved and analyzed by the DSSs to provide decision support solution. The FHIR server allows standardized transfer of data and interoperability between DESIMS and the three DSSs. The standard FHIR resources, namely patient, observation, condition, bodysite, specimen, activities, and careplan are used for input/output data exchange between DESIMS and the CDSSs.

C. Clinical Decision Support Systems

As the aim of DESIREE is to facilitate personalized, collaborative, and multidisciplinary management of primary breast cancer, in the proposed model, in addition to GL-DSS, we integrate the EX and CB-DSS. Fig. 2 shows the workflow and integration of the three DSSs.

In the proposed model, GL-DSS takes a pure evidence-based approach, and is built by referring to some of the recommended standard oncology clinical practice guidelines (CPGs), including NCCN [7] and ESMO [8], but also local protocols from different hospitals are implemented, such as

AP-HP [9], and ONK [10] CPGs. However, as GL-based DSS could suffer from the knowledge gap in CPGs, EXP-DSS explores the previously solved patient cases to model the clinical know-how expressed in non-compliant decisions. It generates EXP-rules for the patient cases, which did not comply with the guidelines. Thereby, by integrating with the GL-DSS, EXP-DSS manages to fill the knowledge gap present in the GL-DSS. Finally, the CB-DSS relies on case-based reasoning (CBR) as a problem-solving paradigm. It considers previous similar cases to provide decision support for the new query case. In the DESIREE platform, it functions independently from GL and EXP-DSS. This gives it the flexibility to incorporate more elements of precision medicine into the model in future, which can then in-turn support the breast cancer screening and treatment more effectively. For example, the availability of genetic data for a higher number of patients in the patient case-base could enable the CB-DSS to identify patients who are susceptible to develop the disease.
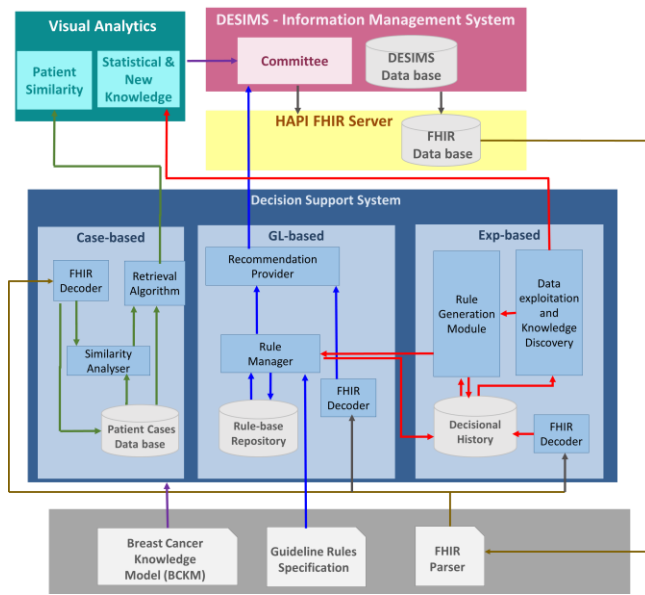

Figure 2. Clinical decision support systems in DESIREE platform

One of the main advantages of the proposed framework is that all the three DSSs are built upon the same breast cancer knowledge model (BCKM), which accounts for the data model and organizes the medical terms in breast cancer oncology based on concepts and relationships between them in an ontological form [11]. The BCKM is represented as an OWL2 ontology. For GL-DSS, the BCKM supports rule-based reasoning as well as subsumption-based reasoning. Whereas the EXP-DSS uses the BCKM to support the rule generation for non-compliance cases. Finally, in CB-DSS, BCKM is employed to weight the semantic similarity measure between patients. In the following, we will describe how each of the DSSs in DESIREE platform performs while integrating with the BCKM.

1) Guideline-based Decision Support System

The aim of the GL-DSS is to deliver patient-specific recommendations for a given patient case on the basis of the selected CPGs represented as sets of IF-THEN rules. To obtain these rule sets, each of the considered CPGs was first described as a set of human-readable decision rules from the original text. Then, these specifications were manually encoded into a formal model-driven rule language (NRL). The formal rules were processed with a Rule Manager web service which checks them against the BCKM and generates an internal computer-interpretable representation of the rules in the N3 notation. Both NRL and N3 rule representations are stored in a rule-base repository.

At execution time, when the Recommendation Provider web service is called, patient data are transferred from the DESIMS using FHIR messages. A FHIR Decoder service transforms the message into a triple representation of the patient data, consistently with the data model and the concepts described in the BCKM. Then, rules are matched to the patient representation using a semantic inference engine (Euler/Eye) [12] that implements rule-based reasoning, according to the knowledge issued from guideline contents and encoded in the rules, and subsumption-based reasoning using domain knowledge organized in the BCKM ontology. Rules, triggered in forward chaining mode, first infer new patient-derived data that enrich the patient raw description and then produce guideline-based recommendations. Recommendations are mainly made of care procedures, grouped or not into care plans applicable to the patient. In some rules, recommendations are provided as atomic entities that can be combined with additional actions or refined by more specific actions to help the reconciliation of recommendations into care plans performed by the Recommendation Provider [13]. Depending on CPG contents, each piece of recommendation has a specific evidence level (as provided in CPGs) and a conformance level categorized as SHALL, SHOULD, MAY, MAYNOT, SHOULDNOT, SHALLNOT. The issued patient-centred recommendations are then returned to the DESIMS and displayed to the MBU users who may decide to comply, or not, with them.

2) Experience-based Decision Support System

The EXP-DSS mainly acts as a supporting mechanism to the GL-DSS. With incorporating the experience of the clinicians, it augments the knowledge contained in the GL-DSS and helps in providing recommendation to certain complex or special patient cases that the CPG is not able to cope with. Those clinical cases are non-compliant with the CPG as clinicians do not follow CPG-based recommendations.

The EXP-DSS presented in Fig. 2, is composed by (i) a rule generation module, (ii) a data exploitation and knowledge discovery module, (iii) a FHIR decoder to retrieve all clinical information in a structured way from the FHIR server and (iv) the decisional history that gathers all

the information related to all decisions made for every patient.

The rule generation module creates new experience-based rules from CPG non-compliant clinical decisions. It receives the response from the BU and analyzes the final decision made by the clinicians. In case the final decision is non-compliant with the CPGs' provided recommendations, a new rule that solves the clinical case is created. The rule is generated by considering (i) the executed rule set and the conditions describing each of these rules, (ii) the non-compliance criteria defined with a single or a set of justifications and (iii) the final decision made by the BU. With further execution of similar clinical cases the recommendations generated by this module will be shown along with the ones coming from the GL-CDSS.

The data exploration and knowledge discovery on the other hand, uses the clinical data retrieved from DESIMS from FHIR messaging for its exploitation by data mining techniques in order to search for correlations among parameters and critical breakpoints for continuous variables related to treatment and outcomes.

Finally, the decisional history, structures all the information related to the decision-making process in a processable way for the EXP-DSS. It contains all the information related with a decision for each patient, which comprises the clinical parameters that characterize the patient, the executed rule set for that patient, the final decision made by the clinicians, the non-compliance criteria when it was a CPG non-compliant case, the executed treatment and the outcomes for evaluating the performance of the decision over time.

3) Case-based Decision Support System

The CB-DSS is based on CBR methodology, which relies on previous experience in form of resolved cases stored in the case base to provide solution to the new problem. A generic CBR system is composed of four consecutive processes, known as the CBR cycle, including retrieval, reuse, revise and retain. In the proposed model, the CB-DSS is built to retrieve similar cases from the case-base, which are then compared in the rainbow boxes visual analytics GUI and displayed to the clinician for decision support.

The similarity retrieval algorithm is defined using "local" and "global" similarity functions. Local similarity function measures the distance between simple attributes in the query case against the ones characterizing patient cases in the case base. The global similarity function uses the results from local similarity measures to compare the compound attributes. In the proposed CBR system, a patient case is represented as a compound attribute, composed of several simple attributes, including physiological and clinical variables. For the simple attributes having categorical values, semantic similarity function is applied to compute the distance between them as a degree of taxonomical proximity. The similarity semantic function, such as the

cosine similarity function shown in Eq. (1), utilizes the hierarchical placement of concepts in the domain ontology in measuring the similarity distance. It measures the similarity between two vectors or sets, thus takes into account of the number of superclasses or ancestors in the ontology.

$$cosine(i_1, i_2) = sim(t(i_1), t(i_2)) = \frac{\left| \left(\bigcup_{d_i \in t(i_1)} (super(d_i, CN)) \right) \cap \left(\bigcup_{d_i \in t(i_2)} (super(d_i, CN)) \right) \right|}{\sqrt{\left| \bigcup_{d_i \in t(t_i)} (super(d_i, CN)) \right|} \cdot \sqrt{\left| \bigcup_{d_i \in t(i_2)} (super(d_i, CN)) \right|}}$$

(1)

Where $super(c, C)$ is the subset of concepts in C which are super concepts of c and $t(i)$ is the set of concepts the individual i is an instance of.

Now, with the local similarity measures computed, the global similarity function is used to select the patient case(s) that are most similar to the query case. The k-Nearest Neighbour (k-NN) method is computed to retrieve the top k similar cases to the query case from the patient case base, according to the global similarity function. Given the query patient case X and the local similarity measure computed for the N patient cases in the case base $Y_j = y_{ji}$, the Euclidian distance between the query and case base is computed. Based on the obtained result, the k nearest patient cases are first located in the case base. The k-NN similarity measure is then computed, which measures the arithmetic mean output across patient cases in the case base and returns a value between 0 ~ 1, with 0 and 1 indicating the retrieved case being less and most similar to the query case, respectively.

D. Visual Analytics GUI

The visual analytics GUI plays a very important part in the DESIREE framework, as it acts as the media through which the results from CB-DSS and EXP-DSS are presented to the clinician. Thus, it is critical that the proposed algorithm properly adapts the results from the DSSs and graphically represent them to the clinician in an intuitive and user-friendly manner.

1) Patient Similarity Visual Analytics

The patient similarity visual analytics aims at supporting the clinicians during the retrieve CBR phase, by providing visual information and arguments. This approach is the exact opposite of the "black box" approach, *i.e.* a system that gives opaque directives such as "you should prescribe this treatment", without explanations.
The visual analytics shows the main attribute values for the query patient and the 'k' most similar patients, and the treatment prescribed to the 'k' most similar patients. Fig. 3 shows an example of the visual analytics. It includes two parts.

The first part (on top in Fig. 3) shows quantitative data, using a scatter plot generated using Multidimensional Scaling (MDS) [14]. The center white dot represents the query patient, and the other, colored, dots represent the similar patients. The assigned color indicates the type of treatment received by the similar patients (red: surgery,

yellow: endocrine therapy, green: chemotherapy, blue: radiotherapy). Distance between the dots represent patient similarity: more similar the patients, closer the dots. In the scatter plot of Fig. 3, one can see that there are three similar patients treated by surgery and nine by endocrine therapy, and that the closest similar patient was treated by surgery.
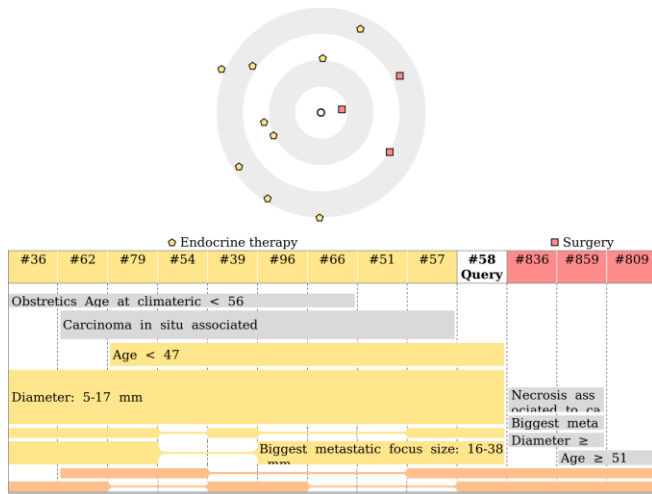

Figure 3. Screenshot of the patient similarity visual analytics.

The second part (bottom of Fig. 3) shows qualitative data, using rainbow boxes [15]. Columns represent retrieved patients, using the pre-defined color code. The query patient is located at the junction of the two main therapeutic options (here surgery and endocrine therapy) which are displayed on the left and the right sides. (Attribute, value) pairs are extracted from the dataset, and the most interesting pairs for taking the therapeutic decision are selected, using Mutual Information (MI; we used MI between the attribute and the treatment). Selected (attribute, value) pairs are displayed as rectangular boxes in rainbow boxes.

Each box covers the columns corresponding to the patients for which the (attribute, value) pair holds. When these patients are not contiguous, "holes" are present in the box (see example for box "Biggest metastatic focus size" in Fig. 3). Columns are ordered using the Artificial Feeding Bird (AFB) metaheuristic [16], in order to minimize the number of holes. Box height is proportional to MI. Grey boxes do not include the query, otherwise, the box color is the mean of the column headers colors. Thus, box color indicates towards which treatment the box orientates.

In Fig. 3, one can see that the main qualitative argument is "diameter between 5-17" (tallest colored box), and that it orientates towards endocrine therapy (yellow color). In fact, the "diameter between 5-17" (attribute, value) pair is shared by the query with all similar patients treated by endocrine therapy, and none of those treated by surgery.

2) Statistical and New Knowledge Visual Analytics

The main goal of the statistical and new knowledge visual analytics is to provide new insights generated by the new management tool in the EXP-DSS. The visualisation tool consists of two different dashboards, namely the statistical and the parallel coordinates dashboards.

As shown in Fig. 4, statistical dashboard consists of six synchronized graphics and it allows the user to select the scenario to display. As numbered in Fig. 4, (1) is the main scatter plot that displays a bubble for each treatment with respect to the two selected variables, (2) treatment type pie chart, (3 & 4) is the bar charts displaying the compliance and non-compliance details, and (5 & 6) represent the toxicity percentage and toxicity types for the selected scenario.

The parallel coordinates dashboard shown in Fig. 5, on the other hand enables discovery of patterns related to clinically relevant outcomes, such as patterns related to treatment response, toxicities or patient outcomes. This interface is designed to be interactive, allowing free exploration of data. This visualization tool lets the user to select the scenario and treatment group to display, and the criteria to color the parallel coordinates lines (usually binary variables). The axes will display new knowledge extracted by the EXP-DSS. The new knowledge visualization module lets the domain experts committee to review the new knowledge generated in the experience based CDSS in an intuitive manner. The visualization system of this module provides the necessary tools to handle and manage the generated new knowledge.
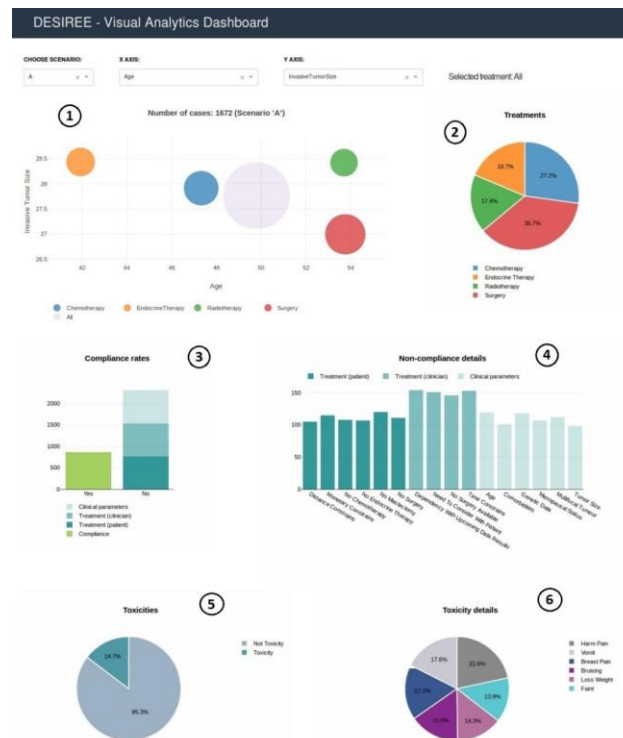
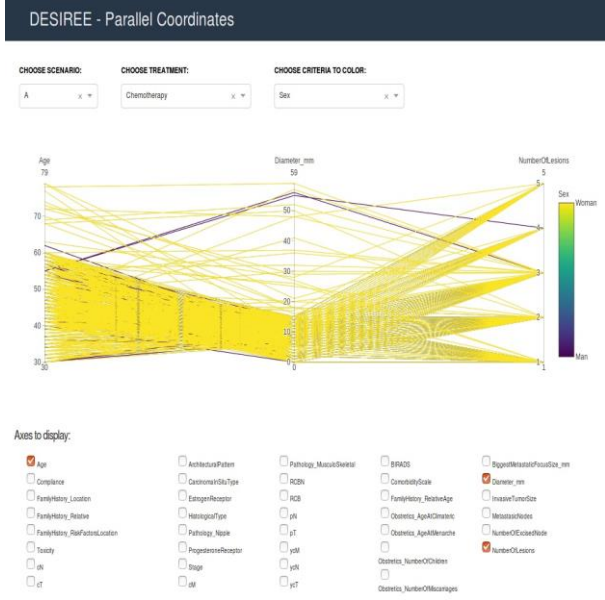

Figure 4. Statistical dashboard

Figure. 5 Parallel coordinates dashboard

III. Discussion and Conclusion

We note two main challenges in the current clinical practice at breast cancer oncology, where MBUs guided by CPGs are introduced to improve the quality of the therapeutic decisions made. MBUs are required to handle large number of patients and are thereby limited on the amount of time dedicated to review each patient. Also, the CPGs used in current practice, can only provide recommendations for "standard" patients, they fail in complex cases. DESIREE project, tries to address these challenges, by developing a web-based software ecosystem for the MBUs. Three CDSSs modalities, namely GL-DSS, EXP-DSS and CB-DSS are developed to improve the therapeutic decisions made in the MBUs. The proposed model, not only reduces the burden on the clinicians in efficiently handling large number of patients, but with EXP-DSS and CB-DSS, it also enables patient-centered healthcare possible.

In this paper, we demonstrate the overall architecture, and workflow of the three DSSs and its integration in the DESIREE platform. Currently, the three modules have been developed and technically validated using either online database or simulated data. First results are presented in this paper based on visual analytics GUI. The next step is to use DESIREE on real patient data, including individual module adaptation with real data, integration of modules, to finally perform the whole clinical validation.

References

[1] GLOBOCAN 2012: Estimated Cancer Incidence, mortality and prevalence worldwide in 2012, by World Health Organization (WHO), from http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx on May 2018.

[2] Breast Cancer rates in under-50s at record high, from NHS choices, https://www.nhs.uk/news/cancer/breast-cancer-rates-in-under-50s-at-record-high/ on May 2018.

[3] B. Séroussi, J. Bouaud, and EC. Antoine, "OncoDoc: a successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer". Artif Intell Med, vol. 22(1), 2001, pp. 43-64.

[4] M.E. Matheny, L. Ohno-Machado, "Generation of Knowledge for Clinical Decision Support: Statistical and Machine Learning Techniques" Clinical Decision Support (2nd Eds.), 2014, pp. 309-337.

[5] T.J. Bright, et. al, "Effect of clinical decision-support systems: a systematic review" Ann Intern Med, vol. 157(1), 2012, pp. 29-43.

[6] N.V. Chawla, and D.A. Davis, "Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework" Journal of General Internal Medicine, vol. 28(3), Sept. 2013, 660-665.

[7] NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) Breast Cancer, Version 1.2016, 02/10/16 © National Comprehensive Cancer Network, Inc. 2016.

[8] E. Senkus, S. Kyriakides, S. Ohno, F. Penault-Llorca, P. Poortmans, E. Rutgers, S. Zackrisson, and F. Cardoso, on behalf of the ESMO Guidelines Committee; Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up, Annals of Oncology, vol. 26, Issue suppl_5, pp v8-v30 September 2015, Pages v8–v30, https://doi.org/10.1093/annonc/mdv298

[9] Référentiel cancer du sein, Assistance Publique–Hôpitaux de Paris, https://www.aphp.fr/file/5462/ [last accessed June 25th, 2018]

[10] Onkologikoa breast cancer guidelines v1 2017. Internal document.

[11] J. Bouaud, G. Guézennec, and B. Séroussi, "Combining the Generic Entity-Attribute-Value Model and Terminological Models into a Common Ontology to Enable Data Integration and Decision Support" Stud Health Technology Inform, vol. 247, 2018, pp. 541-545.

[12] R. Verborgh, and J. De Roo, "Drawing conclusions from linked data on the web: The EYE reasoned", IEEE Software, vol. 32(3), 2015, pp. 23–27.

[13] B. Séroussi, G. Guézennec, J.B. Lamy, N.I. Muro, N. Larburu, B. Sekar, C. Prebet, J. Bouaud, "Reconciliation of multiple guidelines for decision support: a case study on the multidisciplinary management of breast cancer within the DESIREE project", Proc.A MIA Annu Symp, 2017, pp. 527-1536.

[14] I. Borg, P.J.F. Groenen, and P. Mair, "Applied multidimensional scaling" Springer 2013.

[15] J.B. Lamy, H. Berthelot, C. Capron, M. Favre, "Rainbow boxes: a new technique for overlapping set visualization and two applications in the biomedical domain", Journal of Visual Language and Computing, vol. 43, 2017, pp. 71-82.

[16] J.B. Lamy, "Artificial Feeding Birds (AFB): a new metaheuristic inspired by the behavior of pigeons", "Advances in nature-inspired computing and applications", Spinger 2018, in press.