Computational modeling of psychological resilience trajectories during breast cancer treatment

Georgios C. Manikis
Computational Biomedicine Laboratory
FORTH-ICS
Heraklion, Greece
gmanikis@ics.forth.gr

Haridimos Kondylakis
Computational Biomedicine Laboratory
FORTH-ICS
Heraklion, Greece
kondylak@ics.forth.gr

Dimitrios G. Katehakis
Computational Biomedicine Laboratory
FORTH-ICS
Heraklion, Greece
katehaki@ics.forth.gr

Ruth Pat-Horenczyk
Hebrew University School of Social Work, Jerusalem, Israel
ruth.pat-horenczyk@mail.huji.ac.il

Konstantina Kourou
Institute of Molecular Biology and Biotechnology
FORTH-IMBB
Ioannina, Greece
konstadina.kourou@gmail.com

Evangelos Karademas
Computational Biomedicine Laboratory
FORTH-ICS & University of Crete
Heraklion, Greece
karademas@uoc.gr

Lefteris Koumakis
Computational Biomedicine Laboratory
FORTH-ICS
Heraklion, Greece
koumakis@ics.forth.gr

Dimitrios I. Fotiadis
Institute of Molecular Biology and Biotechnology
FORTH-IMBB, Ioannina, Greece
dimitris.fotiadis30@gmail.com

Panagiotis Simos
Computational Biomedicine Laboratory
FORTH-ICS & University of Crete
Heraklion, Greece
akis.simos@gmail.com

Paula Poikonen-Saksela
Helsinki University Hospital
Comprehensive Cancer Center
Helsinki, Finland
paula.poikonen-saksela@hus.fi

Kostas Marias
Computational Biomedicine Laboratory
FORTH-ICS & Hellenic Mediterranean University
Heraklion, Greece
kmarias@gmail.com

Angelina Kouroubali
Computational Biomedicine Laboratory
FORTH-ICS
Heraklion, Greece
kouroump@ics.forth.gr

Manolis Tsiknakis
Computational Biomedicine Laboratory
FORTH-ICS & Hellenic Mediterranean University
Heraklion, Greece
tsiknaki@ics.forth.gr

***Abstract*— Coping with breast cancer and its consequences has now become a major socioeconomic challenge. The BOUNCE EU H2020 project aims at building a quantitative mathematical model of factors associated with optimal adjustment capacity to cancer. This paper gives an overview of the project targets and on the algorithmic methods focusing on modeling the psychological resilience trajectories during breast cancer treatment.**

Keywords*— *Computational Modeling, Machine Learning, Resilience, Cancer

I. Introduction

Breast cancer is the most common cancer in women world-wide accounting for 28% of the total cancer cases in the WHO European Region, while the incidence of breast cancer in the eastern and southern European countries is still rising[1]. Still, there is a very high survival rate among breast cancer patients[2]. Therefore and in order to ensuring a better and faster recovery and higher quality of life, coping with breast cancer and its consequences becomes a major socioeconomic challenge calling for novel strategies for understanding, predicting, increasing the resilience of patients to all the stressful challenges and experiences [1], [2] and providing appropriate recommendations [3], [4], [5].

In this context, an interdisciplinary consortium of experts was formed in 2017 in response to a HORIZON 2020 call, for Personalized Medicine research and innovation solutions, in order to work towards predicting effective adaptation to breast cancer to help women to BOUNCE Back. The broad and general objective of the BOUNCE project[3] is to build a quantitative mathematical model of factors associated with optimal adjustment capacity to cancer. Special emphasis is given to modifiable factors associated with optimal disease outcomes. At the core of the project is a prospective multi-centre clinical pilot at four major oncology centres: European Institute of Oncology in Italy, Helsinki University Hospital in Finland, the Rabin Medical Center, Shaare Zedek Medical Center coordinated by the Hebrew University of Jerusalem in Israel and Champalimaud Foundation in Portugal. The study is currently recruiting over 660 breast cancer patients with stage I-III histologically confirmed diagnosis. There will be seven assessment waves over a period of 18 months: baseline, which will occur after the first visit to the oncologist, Month 3, 6, 9, 12, 15, and 18.

[1] http://www.euro.who.int/en/health-topics/noncommunicable-diseases/cancer/news/news/2012/2/early-detection-of-common-cancers/breast-cancer

[2] https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/survival

[3] https://www.bounce-project.eu

2471-7819/19/$31.00 ©2019 IEEE
DOI 10.1109/BIBE.2019.00082

The primary end-point of the BOUNCE clinical pilot is to identify factors and processes that may predict both interim and long-term patient resilience, their physical wellbeing and psychological outcomes of cancer and cancer treatment. The design of the study will allow researchers to apply statistical models to understand complex interactions of variables both within and across measured domains (e.g., from personality traits to health outcomes, through health-related beliefs and behaviour, with socio-demographic and cultural variables as moderating conditions) in order to gain an enhanced understanding of the dynamic process of adaptation to breast cancer, and resilience-as-a-process.

The rest of the paper is structured as follows: In section 2, we define resilience whereas in Section 3 we present the algorithms that are going to be eventually used for achieving individualized models of resilience. Finally, Section 4 concludes this paper and presents the next steps to be implemented until the end of the project.

II. Defining Resilience

Being resilient does not mean that a person doesn't experience difficulty or distress [6][7]. Emotional pain and sadness are part of the common experience of people who have suffered major difficult life events and trauma [8]. This is also true for breast cancer patients for whom resilience is a major asset for achieving better adaptation to illness and higher quality of life [9][10].

Some understand resilience to be a predisposition or an existing potential, before even facing an adverse situation [11]. In this sense, resilience as a capacity or trait is the integration of the internal and external resources (e.g., optimism, humor, social support) available to the individual facing adversity which may influence/increase the effectiveness of the coping process. Resilience-as-trait measurement is subject to significant drawbacks, for instance people understand themselves in different ways. However, instead of looking at resilience through the “eyes” of the patients--which is subject to report bias and it is also likely to be affected by illness representations, coping strategies etc.--we also look at resilience by modelling the person’s affective and behavioural responses to the disease and to subsequent negative events (i.e. stressors) (Fig. 1). In the present work we focus on analysis techniques addressing the evolution of resilience-as-process which is inferred from the observation of positive adaptation to illness and better outcomes, despite any negative events [11], such as initial diagnosis, subsequent therapy side-effects, negative test results etc. The outcomes at all time points T1-Tn (see Fig. 1) could be on a single dimension (i.e. Quality of Life (QoL) and mental health/affective state and functionality) or complementary outcomes considered separately (i.e. QoL, mental health/affective state, functionality and physical health).

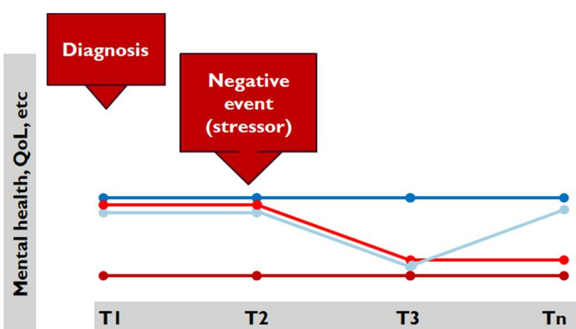

Fig. 1. Schematic representation of trajectories of psychosocial and functional illness outcomes during the course of illness. T1 represents measurements at the time of cancer diagnosis. Within BOUNCE Tn represents each subsequent outcome measurements up to 18 months after diagnosis. Each line indicates one patient or subgroup of patients displaying similar adaptation to illness. Additional stressors may emerge at various time points that may affect the adaptation process in varying degrees.

III. Using Adaptive Algorithms to Achieve Individualized Models of Resilience

Modelling of psychological resilience requires building theoretically plausible, clinically useful, and computationally sound schemes describing: (i) the predominant mechanisms involved in the process of psychological adaptation to cancer and (ii) the most powerful longitudinal predictors of long-term psychosocial and functional outcomes following treatment for breast cancer. BOUNCE aspires to go further than conventional multivariate statistical methods permit and develop a prediction tool that can be used at any point during breast cancer diagnosis and treatment to identify patients at risk for poor psychosocial and functional outcomes—i.e., patients who, at a given point in time, demonstrate poor psychological resilience. Given the inherent complexity of the longitudinal data, BOUNCE will develop and evaluate a Machine Learning (ML) framework to identify subgroups of patients that display distinct psychosocial profiles (at specific time points and over time) in adapting to breast cancer. Firstly, patient clustering is applied to group patients at a given time point according to a specific ‘level’ of adaptation to illness. In this approach resilience is defined according to the observation of affective and functional status. Secondly, longitudinal data are exploited through a clustering methodology aiming to distinguish patient profiles according to possible transitions from one resilience category to another due to changes in specific factors. Finally, all clinical predictive outcomes will be entered into a decision-level fusion model to investigate whether the ensemble of the decisions further improves prediction of resilience at a specific time point. The BOUNCE trajectory predictor will exploit effectively factors measured in the multicenter clinical study. This set of factors consists of: (i) patient-reported outcomes (i.e. mental health, distress level, health- and overall QoL, and functionality), (ii) illness-related self-regulation variables (i.e. self-rated health etc.), (iii) potentially stressful events taking place during the follow up period, (iv) moderators and facilitators (i.e. self-efficacy, resilience, social support etc.) and (v) lifestyle factors (i.e. health habits etc.).

All aforementioned data are collected using standardized questionnaires and are available through the Noona[4] tool. Then data are exported in batches and are continuously integrated using a novel architecture [12] and data infrastructure [13] combining a data lake for staging the available information and an ontology for the integration and their harmonization and a research supporting tool for facilitating data exploration and visualization [14], [15].

A. Unsupervised Machine Learning Techniques

BOUNCE unsupervised learning analysis framework will be coupled with machine learning and conventional

[4] http://www.noona.com/

statistical approaches with the aim to group individual trajectories into distinct clusters and reveal intra-individual changes and inter-individual differences in the examined patients. To this extend, clustering techniques will be also employed to analyze data under a cross-sectional perspective using corresponding data acquired from distinct time points during the critical 18-month period following cancer diagnosis.

Within the BOUNCE cross sectional unsupervised learning framework, symptom clusters will be generated based on current mental health and illness-related distress, QoL, and functional level. Cut-off thresholds will be set to define discrete levels of each condition (e.g. high, moderate, mild and low level of each condition) and clustering performance will be validated both externally, using current patients' condition, and internally based on validity indices such as Silhouette, Dunn, etc. Descriptive statistics and frequency distributions will assess potential statistical differences between the clinical, psychosocial, and behavioural characteristics grouped into the distinct clusters. Illustrative representation of the reported clusters and distributions of all examined data will be provided via clustergrams, radar plots, cluster-change membership, and alluvial diagrams. Robust and sparse k-means clustering (RSKC) [12] will be the model of choice since it is able to: a) deal with missing data and outliers, and b) assign weights to the examined data as they might have varying effects on clustering and not contribute equally in determining the clusters.

The aim of longitudinal clustering is to establish patient profiles across time-points using unsupervised learning (clustering) techniques. Cluster-change profile of each person over time will be also determined revealing the patient's' potential shifting from one adaptation/resilience category to another due to changes in medical and/or psychological and behavioural factors. Within this analysis framework persons who belong to cluster A at T1, Cluster B at T2 and Cluster A again at T3 will be grouped together. Determining the shifting from one category to another during the trajectory of the disease will enable the estimation of the exact number of patients that belong to each category. Moreover, an accurate estimation for the possibility of shifting from low to high resilience and vice versa across different time points will be achieved. The overall process of adaptation to illness and the resilience level will be grouped into certain clusters of patients' characteristics and behaviour; thus, empowering the prediction of final outcomes according to the categories that are most informative across time. Latent growth curve modelling (LGCM) and growth mixture modelling (GMM) have frequently been employed to handle data on disease progress of oncological patients after treatment, showing interesting results in trajectory clustering and identifying behavioural risk factors as predictors of trajectory groups [17]. A machine learning based resilience clustering approach will be also applied providing distinct clusters not only on the basis of individual trajectory similarities across time but on the trajectory shape as well. K-Means for longitudinal data using shape-respecting distance has recently been investigated demonstrating higher performance when compared to traditional longitudinal clustering techniques [18].

Within this analysis, the factors and/or their interactions that can accurately predict final and intermediate outcomes will be identified. The medical and psychological/ behavioural factors that will be considered in the supervised analysis framework will be assessed at: i) previous time interval, ii) at baseline and previous time interval and iii) across different time intervals within the 18-month follow up period. Towards this direction, the development of the final BOUNCE predictive tool will be achieved by exploiting factors over time.

B. *Supervised Machine Learning Techniques*

Supervised machine learning and conventional statistical methods will be adopted to further exploit and model the heterogeneous multiscale data within BOUNCE both longitudinally and at single time intervals. This type of analysis will enable the prediction of intermediate and final outcomes related to illness adaptation and resilience. The different patients' profiles in terms of the scores in the medical and psychological/behavioural variables assessed during the follow-up period will be considered for prediction purposes (Fig.2).

More specifically, concerning the cross-sectional studies within the BOUNCE supervised learning analysis framework, several predictive models (extreme gradient boosting-XGBoost, generalized linear models-GLM, random forests-RF, weighted random support vector machine clusters analysis-WRSVMC, etc.) will be trained, validated and tested under a nested cross-validation schema and model comparison will be reported using exactly the same input data during all iterations. Feature ranking/selection mechanisms will be implemented during model training. A probabilistic outcome will be provided related to resilience status and all medical, psychological, and behavioral factors will be ranked according to their importance in the predictive performance. Model performance will be quantified using several statistical metrics including accuracy, sensitivity or recall, specificity, precision, F1-score and area under the curve (AUC).

BOUNCE longitudinal analysis will include, among other models, a novel semi-parametric marginal approach (boosted multivariate trees for longitudinal data-boostmtree) as reported in [15] to identify all related interactions between BOUNCE medical, psychological, and behavioural factors and time semi-nonparametrically. The most important factors and factor-time interactions will be identified using permutation variable importance techniques. Growth-based trajectory modelling will be used to classify patients according to their adaptation/resilience level at final (i.e., at 18 months) and intermediate outcomes (i.e., at 6, 12… months). Several regression models will fit BOUNCE longitudinal data simultaneously and patient-specific probability of group membership will be assigned. Additionally, group-based trajectories will be estimated for each group of patients over time and goodness of fit accuracy will be assessed using C-statistics and Bayesian information criterion (BIC).

C. *Design of the Predictor Model Aggregation*

Within BOUNCE, the prediction tool includes a limited number of biomedical factors and self-ratings, that will emerge (and be validated) as significant predictors of outcomes in addition to anxiety, depression, and distress which are measured early in the course of the disease. Focusing on end-point outcomes instead of trajectories the predictive outcome will be continuous or categorical. To this

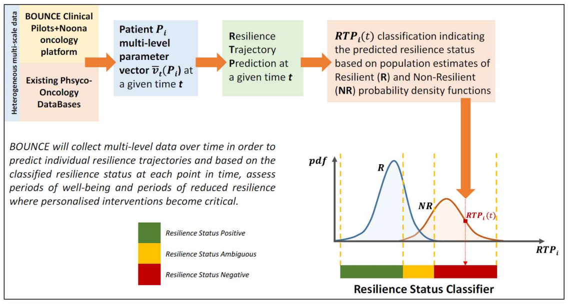

Fig. 2. BOUNCE cross-sectional predictive modelling framework for resilience status.

end, good outcomes at Tn regardless of prior status imply indirectly high resilience, whereas poor outcomes imply low resilience as has been defined in BOUNCE.

The fusion model within BOUNCE is implemented by utilizing ensemble methods which combine the predictions of several base estimators. The base estimators are built with given learning algorithms, such as generalized linear models, random forests, extreme gradient boosting, support vector machines-SVM, etc., in order to improve the performance compared to a single estimator. Voting classifier [20] is applied for combining conceptually different machine learning classifiers and use a majority vote ("hard" vote) or the average predicted probabilities ("soft" vote) to predict the class labels. Such a classifier can be useful for a set of equally well performing supervised or unsupervised models in order to balance out their individual weaknesses. With the ensemble vote classifier different training sets are considered for building the predictive models (Fig.3). As mentioned above, different classification algorithms are adopted for predicting the class label of new samples. Voting classifier will be applied for making the final and more robust prediction of the end-point outcomes.

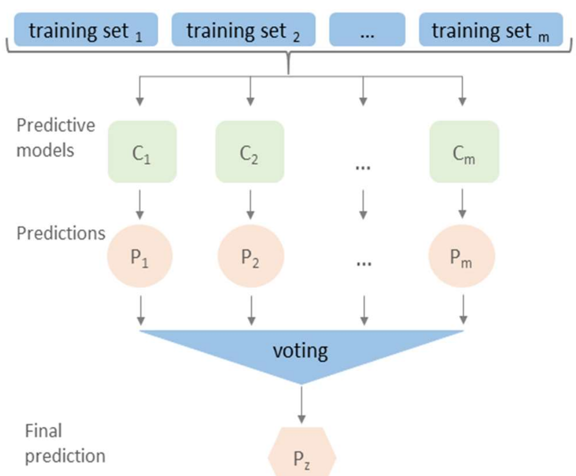

Fig. 3. The process followed by an ensemble voting classifier for combining the different predictions and make the final prediction in terms of voting. Different training sets are considered for each classifier in order to build the predictive models before their fusion.

In sum, BOUNCE develops and deploys advanced computational tools to validate indices of patients' capacity to bounce back during the highly stressful treatment and recovery period following diagnosis of breast cancer. Elements of a dynamic, predictive model of patient outcomes are incorporated in building a decision-support system to be used in routine clinical practice providing oncologists and other health professionals with concrete, personalized recommendations regarding optimal psychosocial support strategies.

IV. Conclusions

This paper presents an overview of the BOUNCE project for developing an individualized model of resilience. This model will be constructed using unsupervised and supervised machine learning techniques that will be combined through model aggregation techniques, in order to gain an enhanced understanding of the dynamic process of adaptation to breast cancer. Currently the project is collecting data, whereas the data infrastructure and the research support tool for data visualization and exploration are already available. The next step is the implementation of the model execution engine and the identification of the specific models to be used.

Acknowledgment

Work presented in the paper is part of the BOUNCE project and has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777167. Any opinions, results, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of BOUNCE or the European Commission.

References

[1] G. Ristevska-Dimitrovska, I. Filov, D. Rajchanovska, P. Stefanovski, B. Dejanova, "Resilience and Quality of Life in Breast Cancer Patients," Open Access Maced. J. Med. Sci. 3, pp.727–731, 2015.

[2] H. Kondylakis, et al. "iManageCancer: Developing a platform for Empowering patients and strengthening self-management in cancer diseases." 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS). IEEE, pp. 755-760, 2017.

[3] H. Kondylakis, et al. "Semantically-enabled Personal Medical Information Recommender." International Semantic Web Conference (Posters & Demos). 2015.

[4] M. Stratigi, H. Kondylakis, K. Stefanidis. "Fairness in group recommendations in the health domain." 2017 IEEE 33rd International Conference on Data Engineering (ICDE). IEEE, 2017.

[5] H. Kondylakis, et al. "Patient Empowerment through Personal Medical Recommendations." MedInfo 216, pp. 1117, 2015.

[6] American Psychological Association. "The Road to Resilience". Retrieved from http://www.apa.org/helpcenter/road-resilience.aspx, visited August 2019.

[7] A.S. Masten, "Ordinary magic: Resilience processes in development," American Psychologist, 56, pp. 227–238, 2001.

[8] G.A. Bonanno, "Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events?" American psychologist, 59(1), pp. 20, 2004.

[9] B. Izydorczyk, A. Kwapniewska, S. Lizinczyk, K. Sitnik-Warchulska, "Psychological Resilience as a Protective Factor for the Body Image in Post-Mastectomy Women with Breast Cancer," International Journal of Environmental Research and Public Health, 15, pp. 1181, 2018.

[10] S.E. Markovitz, W. Schrooten, A. Arntz, M.L. Peters, "Resilience as a predictor for emotional response to the diagnosis and surgery in breast cancer patients," Psycho-Oncology, 24, pp. 1639-1645, 2015.

[11] S.M. Southwick, G.A. Bonanno, A.S. Masten, C. Panter-Brick, R. Yehuda, "Resilience definitions, theory, and challenges: Interdisciplinary perspectives,". European journal of Psychotraumatology, 5, 25338, 2014.

[12] K. Kourou, H. Kondylakis, L. Koumakis, G.C. Manikis, K. Marias, M. Tsiknakis, P.G. Simos, E. Karademas, D.I. Fotiadis, "A Reference Architecture for Predicting Resilience Levels of Women with Breast

Cancer," IEEE International Conference on Biomedical and Health Informatics (BHI'19) , 2019.

[13] H. Kondylakis, L. Koumakis, D. Katehakis, A. Kouroubali, K. Marias, M. Tsiknakis, P. Simos, E. Karademas, "Developing a Data Infrastructure for Enabling Breast Cancer Women to BOUNCE Back", IEEE CBMS, pp. 652-657, 2019.

[14] K. Kourou, H. Kondylakis, L. Koumakis, G.C. Manikis, K. Marias, M. Tsiknakis, P.G. Simos, E. Karademas, D.I. Fotiadis, "Computational Models for Predicting Resilience Levels of Women with Breast Cancer," MEDICON, 2019.

[15] Koumakis, L., Kondylakis, H., Chatzimina, M., Iatraki, G., Argyropaidas, P., Kazantzaki, E., Tsiknakis, M., Kiefer, S. and Marias, K., "Designing smart analytical data services for a personal health framework". pHealth, 224, pp.123-128, 2016.

[16] Y. Kondo, M. Salibian-Barrera and R. Zamar. "RSKC: An R Package for a Robust and Sparse K-Means Clustering Algorithm", Journal of Statistical Software, 72(5), 1 – 26, 2016.

[17] Y. Yang, J. Cameron, C. Bedi, and G. Humphris, "Fear of cancer recurrence trajectory during radiation treatment and follow-up into survivorship of patients with breast cancer," BMC cancer, vol. 18, p. 1002, 2018.

[18] Genolini, C., Ecochard, R., Benghezal, M., Driss, T., Andrieu, S., & Subtil, F., "kmlShape: an efficient method to cluster longitudinal data (time-series) according to their shapes". Plos one, 11(6), e0150738, 2016.

[19] A. Pande, L. Li, J. Rajeswaran, J. Ehrlinger, U. B. Kogalur, E. H. Blackstone, et al., "Boosted multivariate trees for longitudinal data. Machine learning", vol. 106, pp. 277-305, 2017

[20] D. Ruta and B., "Gabrys: Classifier selection for majority voting. Information fusion", vol. 6, pp. 63-81, 2005