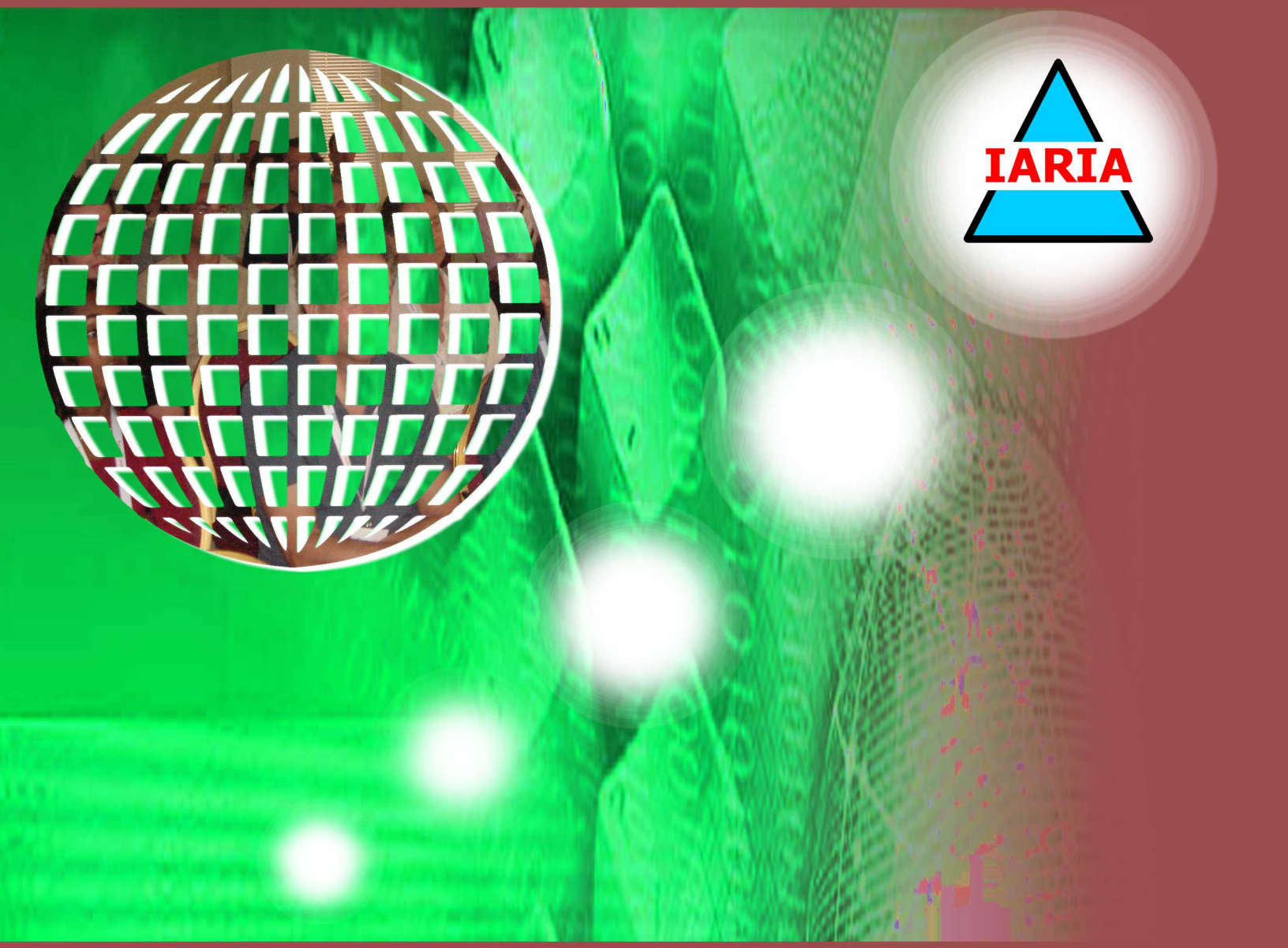
International Journal on Advances in Life Sciences



2019 vol. 11 nr. 1&2

The *International Journal on Advances in Life Sciences* is published by IARIA.
ISSN: 1942-2660
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Life Sciences
Volume 11, Number 1 & 2, 2019

Valentinas Klevas, Kaunas University of Technology / Lithuaniain Energy Institute, Lithuania
Anant R Koppar, PET Research Center / KTwo technology Solutions, India
Bernd Krämer, FernUniversität in Hagen, Germany
Ramesh Krishnamurthy, Health Systems and Innovation Cluster, World Health Organization - Geneva, Switzerland
Roger Mailler, University of Tulsa, USA
Dirk Malzahn, OrgaTech GmbH / Hamburg Open University, Germany
Salah H. Mandil, eStrategies & eHealth for WHO and ITU - Geneva, Switzerland
Herwig Mannaert, University of Antwerp, Belgium
Agostino Marengo, University of Bari, Italy
Igor V. Maslov, EvoCo, Inc., Japan
Ali Masoudi-Nejad, University of Tehran , Iran
Cezary Mazurek, Poznan Supercomputing and Networking Center, Poland
Teresa Meneu, Univ. Politécnica de Valencia, Spain
Kalogiannakis Michail, University of Crete, Greece
José Manuel Molina López, Universidad Carlos III de Madrid, Spain
Karsten Morisse, University of Applied Sciences Osnabrück, Germany
Ali Mostafaeipour, Industrial engineering Department, Yazd University, Yazd, Iran
Katarzyna Musial, King's College London, UK
Hasan Ogul, Baskent University - Ankara, Turkey
José Luis Oliveira, University of Aveiro, Portugal
Hans C. Ossebaard, National Institute for Public Health and the Environment - Bilthoven, The Netherlands
Carlos-Andrés Peña, University of Applied Sciences of Western Switzerland, Switzerland
Tamara Powell, Kennesaw State University, USA
Cédric Pruski, CR SANTEC - Centre de Recherche Public Henri Tudor, Luxembourg
Andry Rakotonirainy, Queensland University of Technology, Australia
Robert Reynolds, Wayne State University, USA
Joel Rodrigues, Institute of Telecommunications / University of Beira Interior, Portugal
Alejandro Rodríguez González, University Carlos III of Madrid, Spain
Nicla Rossini, Université du Luxembourg / Università del Piemonte Orientale / Università di Pavia, Italy
Addisson Salazar, Universidad Politecnica de Valencia, Spain
Abdel-Badeeh Salem, Ain Shams University, Egypt
Matthieu-P. Schapranow, Hasso Plattner Institute, Germany
Åsa Smedberg, Stockholm University, Sweden
Chitsutha Soomlek, University of Regina, Canada
Monika Steinberg, University of Applied Sciences and Arts Hanover, Germany
Jacqui Taylor, Bournemouth University, UK
Andrea Valente, University of Southern Denmark, Denmark
Jan Martijn van der Werf, Utrecht University, The Netherlands
Liezl van Dyk, Stellenbosch University, South Africa
Lisette van Gemert-Pijnen, University of Twente, The Netherlands
Sofie Van Hoecke, Ghent University, Belgium
Iraklis Varlamis, Harokopio University of Athens, Greece
Genny Villa, Université de Montréal, Canada
Stephen White, University of Huddersfield, UK
Levent Yilmaz, Auburn University, USA
Eiko Yoneki, University of Cambridge, UK

International Journal on Advances in Life Sciences
Volume 11, Numbers 1 & 2, 2019

CONTENTS

pages: 1 - 12
Home-based automated assessment of upper limb motor function in Parkinson's Disease
Roberto Nerino, Institute of Electronics, Computer and Telecommunication - National Research Council, Italy
Claudia Ferraris, Institute of Electronics, Computer and Telecommunication - National Research Council, Italy
Giuseppe Pettiti, Institute of Electronics, Computer and Telecommunication - National Research Council, Italy
Antonio Chimienti, Institute of Electronics, Computer and Telecommunication - National Research Council, Italy
Corrado Azzaro, Istituto Auxologico Italiano, Italy
Giovanni Albani, Istituto Auxologico Italiano, Italy
Lorenzo Priano, Istituto Auxologico Italiano, Italy
Alessandro Mauro, Istituto Auxologico Italiano, Italy

pages: 13 - 22
The Role of Adaptive Immersive Technology in Creating Personalised Environments for Emotional Connection and Preservation of Identity in Dementia Care: Insights from User Perspectives towards SENSE-GARDEN
Gemma Goodall, NTNU/Norwegian University of Science and Technology, Norway
Ileana Ciobanu, University of Medicine and Pharmacy Carol Davila, Bucharest, ELIAS University Hospital, Romania
Ronny Broekx, Innovation Department, ePoint, Belgium
Jon Sørgaard, NTNU/Norwegian University of Science and Technology, Norway
Iulian Anghelache, Research & Development Department, Compexin SA, Romania
Catalina Anghelache-Tutulan, Research & Development Department, Compexin SA, Romania
Mara Diaconu, Research & Development Department, Compexin SA, Romania
Sigrid Mæland, Odda Municipality, Norway
Therese Borve, Odda Municipality, Norway
Audun Digranes Dagestad, Odda Municipality, Norway
Piet Bormans, Aan de Beverdijk, Vulpia Vlaanderen, Belgium
Marleen Custers, Aan de Beverdijk, Vulpia Vlaanderen, Belgium
Katrin Losleben, Center for Women's and Gender Research, UiT The Arctic University of Norway, Norway
Rita Valadas, Santa Casa da Misericórdia de Lisboa, Portugal
Cristina Vaz de Almeida, Santa Casa da Misericórdia de Lisboa, Portugal
Alda Matias, Santa Casa da Misericórdia de Lisboa, Portugal
Andreea Marin, University of Medicine and Pharmacy Carol Davila, Bucharest, ELIAS University Hospital, Romania
Kristin Taraldsen, NTNU/Norwegian University of Science and Technology, Norway
Walter Maetzler, 9Department of Neurology, Christian-Albrechts-University, Kiel, Germany
J. Artur Serrano, NTNU/Norwegian University of Science and Technology, Norway

pages: 23 - 32
LASSO Regression for Monitoring Patients Progress Following ACL Reconstruction via Motion Sensors: A Case Study
Salvatore Tedesco, Tyndall National Institute, University College Cork, Ireland
Brendan O'Flynn, Tyndall National Institute, University College Cork, Ireland

pages: 33 - 43
Knowledge Distillation from Machine Learning Models for Prediction of Hemodialysis Outcomes
Harry Freitas da Cruz, Hasso Plattner Insitute, Germany
Siegfried Horschig, Hasso Plattner Insitute, Germany

Christian Nusshag, Heidelberg University Hospital, Germany
Matthieu-P. Schapranow, Hasso Plattner Insitute, Germany

pages: 44 - 55
Identification of Appropriate Facial Expressions for Medical Doctors
Akemi Oe, Graduate School of Software and Information Science Iwate Prefectural University, Japan
Yoshitoshi Murata, Graduate School of Software and Information Science Iwate Prefectural University, Japan

pages: 56 - 67
Sensitive Data Discovery in Care Pathways using Business Process Modelling and HL7-CDA
Intidhar Essefi, University of Tunis el Manar, The Higher Institute of Medical Technologies of Tunis Research Laboratory of Biophysics and Medical Technologies, Tunis, Tunisia
Hanene Boussi Rahmouni, University of the West of England, the Computer Science Research Center, Bristol, UK
Mohamed Fethi Ladeb, Radiology Department Kassab Orthopedics Institute, Manouba, Tunisia

pages: 68 - 74
Strength Training – induced Left Ventricular Remodeling in Heart Failure Patients Short Paper
Zahra Sadek, Laboratory EA 3300 (APERE) « Adaptations Physiologiques à l'Exercice et Réadaptation à l'Effort », Picardie de Jule Vernes University, Lebanon
Said Ahmaidi, Laboratory EA 3300 (APERE) « Adaptations Physiologiques à l'Exercice et Réadaptation à l'Effort », Picardie de Jule Vernes University, France
Mahmoud Youness, Medical Research Center Beirut Cardiac Institute Beirut, Lebanon, Lebanon
Pierre Nassar, Medical Research Center Beirut Cardiac Institute Beirut, Lebanon, Lebanon
Righab Hamdan, Medical Research Center Beirut Cardiac Institute Beirut, Lebanon, Lebanon
Wissam Joumaa, Laboratoire Rammal Hassan Rammal, équipe de recherche PhyToxE, Faculté des Sciences (section V), Université libanaise, Nabatieh, Lebanon, Lebanon
Ali Salami, Laboratoire Rammal Hassan Rammal, équipe de recherche PhyToxE, Faculté des Sciences (section V), Université libanaise, Nabatieh, Lebanon, Lebanon
Wiam Ramadan, Lebanese Institutes for Biomedical Research and Application (LIBRA) International University of Beirut (BIU) and Lebanese International University (LIU) Beirut, Lebanon, Lebanon

pages: 75 - 84
Wrist Actigraphy Analysis From Motionwatch8 Data
Gianluigi Delucca, Medricerca di Gianluigi Delucca, Italy

pages: 85 - 98
Enforcing Genetic Consent and Restrictions through a Privacy-Focused Ontology
Michael Reep, George Mason University, US
Bo Yu, George Mason University, US
Duminda Wijesekera, George Mason University, US
Paulo Costa, George Mason University, US

Home-Based Automated Assessment of Upper Limb Motor Function in Parkinson's Disease

Roberto Nerino, Claudia Ferraris, Giuseppe Pettiti, Antonio Chimienti
Institute of Electronics, Computer and Telecommunication
National Research Council
Turin, Italy
e-mail: roberto.nerino@ieiit.cnr.it, claudia.ferraris@ieiit.cnr.it, giuseppe.pettiti@ieiit.cnr.it, antonio.chimienti@ieiit.cnr.it

Corrado Azzaro, Giovanni Albani, Lorenzo Priano, Alessandro Mauro
Department of Neurology and Neurorehabilitation
Istituto Auxologico Italiano, IRCSS
Piancavallo, Verbania, Italy
e-mail: c.azzaro@auxologico.it, g.albani@auxologico.it, lorenzo.priano@unito.it, mauro@auxologico.it

***Abstract*— This work presents a non-invasive low-cost system suitable for the at home assessment of the neurological impairment of patients affected by Parkinson's Disease (PD). The assessment is automatic and it is based on the accurate tracking of hands and fingers movements of the patient during the execution of standard upper limb tasks specified by the Unified Parkinson's Disease Rating Scale (UPDRS). The system is based on a human computer interface made by light gloves and an optical tracking RGB-Depth device. The accurate tracking and characterization of hands and fingers movements allows both the automatic and objective assessment of UPDRS tasks and the gesture-based management of the system, making it suitable for motor impaired users, as are PD patients. The assessment of UPDRS tasks is performed by a machine learning approach, which uses the kinematic parameters that characterize the patient movements, as input to trained classifiers, with the aim of automatically rating the UPDRS scores of the performance. The classifiers have been trained by an experimental campaign, where cohorts of PD patients were contemporary assessed by a neurologist and the system. Results on the accuracy of the system assessments, as compared to the neurologist's ones, are given, along with preliminary results on monitoring experiments at home. Details about the user interfaces of the system, specifically designed for home-monitoring, are provided. The clinimetric properties of the system and its usability have been evaluated and reported. The results confirm that the system is suitable for the remote monitoring of PD patients at-home.**

Keywords - Parkinson's disease; UPDRS assessment; RGB-D camera; human computer interface; tele-monitoring.

I. INTRODUCTION

This article is an extended version of the paper presented at the Fourth International Conference on Smart Portable, Wearable, Implantable and Disability-oriented Devices and Systems, SPWID 2018, and in particular at the PARKTECHNO special track, where some studies on new technologies for people with Parkinson's Disease were presented [1]. Parkinson's Disease (PD) is a chronic neurodegenerative disease characterized by a progressive impairment in motor functions (e.g., bradykinesia) [2], with important negative impacts on the quality of life. The Unified Parkinson's Disease Rating Scale (UPDRS) [3] is an international evaluation scale, commonly used by neurologists to assess the severity of the disease, whose motor symptoms are the most important and characterizing aspect. Specifically, standardized motor tasks, described into the Section III of the UPDRS and dedicated to the motor examination, are used by neurologists to assess impairments and to assign a subjective score, for each task, on a scale of five classes of increasing severity, from 0 (no impairment) to 4 (severe impairment).

The assessment process takes into account specific kinematic features of the movements (such as amplitude, speed, rhythm variations) and anomalies (such as hesitations, freezing, incomplete movements), which are qualitatively and subjectively evaluated by neurologists. On the other hand, a quantitative and objective assessment of these tasks is considered important to increase the reliability of the clinical assessment [4] and to support the disease management and the patient care. A commonly adopted solution is to make use of the well-established correlation existing between kinematic parameters of the movements and the severity of the impairment [5][6]. This correlation is used in the automatic and objective assessment of UPDRS motor tasks by several technological approaches, including those based on optical devices and wearable inertial sensors [7][8].

Another aspect to be considered is that drug treatment of the PD symptoms is crucial to reduce the effects of the impairment in daily activities. Because of possible fluctuations in impairment, it would be desirable to adjust the therapy on a weekly basis, both for the best effectiveness of the therapy and to reduce the side and long-term effects [9]. Unfortunately, the cost of a traditional weekly assessment, preferably at home to reduce patient's discomfort, is unsustainable for the health care system. In this context, technology can support neurologists with an objective and quantitative assessment of the UPDRS motor tasks.

The paper is organized as follows. The state of the art on the technological approaches adopted in the analysis of the upper limb movements during UPDRS tasks is presented in Section II. The methodological approach we propose for the accurate tracking of hand and fingers movement is described

in Section III. In the same section, the graphical user interfaces, the methods used to evaluate the system usability and the agreement between standard and system assessments are also described. In Section IV, we present the results on the kinematic parameters selection, the automatic classification of the motor performance and the usability of the system. Furthermore, preliminary data about the assessment of patient's performance at home are provided. Conclusions and future work are discussed in Section V.

II. STATE OF THE ART

Several solutions have been proposed for the characterization of upper limb movements during the execution of UPDRS tasks. Approaches based on wireless inertial measurement devices (such as accelerometers and gyroscopes) [9][10][11] and on resistive bending sensors [12] do not suffer of occlusion problems, but they are more uncomfortable for people with mobility difficulties compared to the optical approaches and, more importantly, their invasiveness can affect motor performance.

Recently, optical approaches have been proposed for the hand tracking and the automated assessment of the upper limb tasks of UPDRS, namely: Finger Tapping (FT), Opening-Closing (OC) and Pronation-Supination of the hand (PS). In particular, solutions have been developed based on RGB cameras [13], passive markers [14] and bare hand tracking by consumer depth sensing devices [15][16][17][18].

Less attention is generally given to the assessment of the tracking accuracy obtainable by the proprietary hand-tracking firmware of these consumer devices. Their accuracy can be unsatisfactory especially for fast movements, as has been shown by comparisons with standard optoelectronic systems [19]. Nevertheless, accuracy is an important requirement to be considered for the reliability of kinematic parameters and the assessment of the motor performance. Furthermore, the short life of these devices and the related Software Development Kit (SDK) warns against solutions that are too dependent on proprietary hardware and software.

Along this line of research, we present a low-cost system for the home-based automated assessment of the three upper limb tasks of the UPDRS (i.e., FT, OC, PS). The system hardware is based on lightweight colored gloves, an RGB-Depth sensor and a monitor, while the software implements the 3D tracking of the hand trajectories, characterizes them by kinematic features and assesses the motor performance by Machine Learning algorithms (i.e., trained supervised classifiers). The software performs the real-time tracking by fusion of both color and depth information from the RGB and depth streams. The system acts at the same time as a non-invasive Human Computer Interface (HCI), which allows PD patients with motor impairments to self-manage the execution of the tests.

Respect to other approaches, based only on depth information and proprietary algorithms, the hand tracking is more robust and accurate for fast movements [19], making the final assessment more reliable. Another important characteristic of our solution is that it does not depend on any particular hardware or SDK; it assumes only the availability of RGB and depth streams at reasonable frame rate. Moreover, the accuracies obtained by the classifiers demonstrate the feasibility of the system in the remote assessment of the upper limb tasks of UPDRS. Some preliminary results are provided on the home monitoring of PD patients.

This version extends the conference paper providing more details about the natural user interface, specifically developed to allow patients the self-management of the tasks execution and the interaction with the system. The main features of the supervising component of the monitoring platform, designed to analyze the patient performance remotely, are presented. Finally, considering the importance of the usability aspects of a technology, results of a Post-Study System Usability Questionnaire (PSSUQ) are also presented.

III. SYSTEMS AND METHODS

A. System Hardware

The hand/fingers tracking hardware consists of a low-cost RGB-Depth device (Intel RealSense SR300 ©) that provides synchronized RGB color and Depth streams at resolutions of 1920x1080 (Full HD) at 30fps and 640x480 (VGA) at 30 fps (max. 200) respectively. The RGB-Depth device is connected via a USB port to a personal computer (PC) running Microsoft Windows and equipped with a monitor positioned in front of the user (Figure 1). The monitor provides the visual feedback of the HCI for the user's hand and finger movements. The user equipment consists of black lightweight gloves with imprinted color markers: each color marker corresponds to a particular part of hand to be tracked (e.g., fingertips and wrist) or to be used for color calibration and system interaction (e.g., palm). The working volume of the system is a pyramid trunk, which extends from 0.5m up to 2m from the RGB-Depth sensor. This guarantees enough space to perform the exercises comfortably.

The device drivers and our developed software are used to implement both the hand/fingers tracking and the HCI

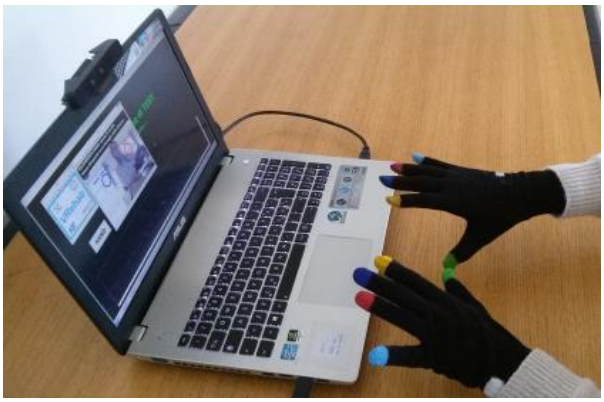


Figure 1. Hand/fingers tracking system

user interface. The software running on the PC implements the acquisition and processing of the data streams for the hand/fingers tracking, the kinematic parameter estimation and the task assessment.

Furthermore, the data produced in every test session, including video sequence of the performance, extracted kinematic parameters and system scores are automatically encrypted and archived for further analysis and for clinician supervision and independent assessment.

B. Human Computer Interface for System Management

The software for the real-time hand/fingers tracking and the graphical user interfaces support the human computer interaction, thanks to which the patient can manage the test session (e.g., start and end the session, select a specific task, enter information on the perceived health status, etc.). Simple gestures, such as opening and closing the hand or pointing the fingers toward the interactive objects of the graphic menu displayed on the monitor, trigger specific actions.

The hand tracking software requires an initial setup phase, which consists of the global adjustment of the image brightness, the detection of the hand area and the color calibration for marker recognition and segmentation. The Intel LibRealSense library is used for the acquisition of RGB and Depth streams, while the OpenCV library [20] is used to retrieve the 3D position of the hand centroid from the Depth stream. A shaking movement of the user's hand starts the recovering of the initial hand position. The hand centroid is used to segment the hand from the background and to define 2D and 3D hand bounding boxes, both for color and depth images. Then, the RGB stream is converted to the HSV color space, more robust to brightness variations.

The design of the color markers and the implementation of a color constancy algorithm compensate for the different lighting conditions that could be found in domestic environments. For this purpose, during the initial setup, the white circular marker on the palm is detected and tracked in the HSV stream. The average levels of each HSV component of the white marker area are used to compensate for the predominant chromatic components due to the different types of lighting. Their values are used to scale each of the three HSV video sub-streams during the tracking phase.

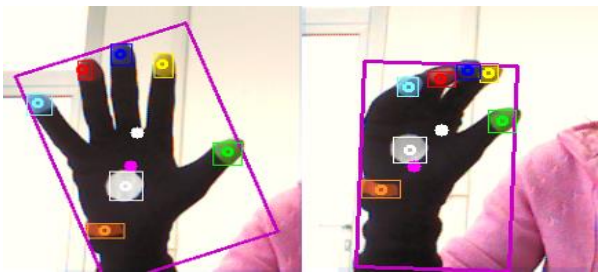


Figure 2. Hand segmentation and marker detection: color blob centroids and bounding box

During the tracking phase, the 3D position of the hand centroid is used to continuously update the 2D and 3D hand bounding boxes (Figure 2). The color thresholds, selected during the initial setup phase, are used to detect and track all the color blobs of the markers. To improve performance and robustness, the CamShift algorithm [20] has been used in the tracking procedure. The 2D pixels of the area of every color marker are reprojected to the corresponding 3D points by standard reprojection algorithms to evaluate the 3D centroid of each color blob. Each centroid is an estimate of the 3D position of the corresponding part of the hand.

The trajectories of all centroids characterize the movements of the hand, which are used to evaluate the task performance (Subsection F).

C. Graphical User Interfaces of the system

The graphical user interfaces (GUIs) of the system become active automatically a few seconds after the system is switched on. The GUIs support two different functionalities, depending on the type of user. The patient GUI is displayed on the monitor at home, and provides the user with visual feedback concerning the movement of the hand and fingers, the actions triggered, and the input given (Figure 3). The GUI menus of the patient interface are managed only by hand gestures, allowing to start/end the session and to select the task to be performed, confirming the choice by closing the hand. Furthermore, the predefined menu items allow the input of some basic information concerning the patient's perceived condition and the type and dosage of drug taken.

Textual messages support the subject during the entire test session and the interaction with the system; in addition, a video guide can be activated by dedicated menu items if the patient has doubts on the correct execution of the task.

Regarding the clinician subsystem, the GUI provides the clinical management and the remote supervision of the patients. The GUI is designed for technical users without disabilities, and consists of a more complex structure, widgets and functionalities. In this case, the GUI input is provided by mouse and keyboard. The GUI is organized as a hierarchy of windows activated by visual objects that trigger the execution of specific actions. Preliminary authentication,

Figure 3. Human computer interface with natural gesture-based interaction of patient subsystem: example of GUI for the task selection

via personal credentials, guarantees a secure access to data only to authorized clinicians.

The main window GUI (Figure 4) allows to select the patient's folder from a repository in which videos, kinematic parameters and system scores of each performance were stored. The clinician can select a particular performance recorded among all those archived; then videos, reports with automated scores and information entered by the patient are displayed to be analyzed by the neurologist. Each video is managed by the functions of a standard video player object (start/stop/pause, rewind, slow motion, etc.), which allows a detailed analysis of the patient's performance. In the lower area of the main window, the clinician can add useful annotations to the record, including the clinical assessments. This information is then stored as part of the patient's record. Messages or communications to patient can be written by the clinician into the dedicated "MSG TO PATIENT" box to be displayed on the main GUI of the patient subsystem before starting the next session. From the main window GUI, four other child windows can be opened.

The first child window (Figure 5), activated from the menu bar of the main window, provides a GUI that is intended to analyze and compare the performances, for the different upper limb tasks, of the left and right hand in the same acquisition session. In the graph area, the radar plots generated by the kinematic parameters of the left and right hand performances are displayed. The average values of the parameters for the UPDRS 0 class (green line), which are estimated from the reference database as described in Subsection D, are also displayed. They are used as reference values for a quick visual comparison of the patient's performance. The UPDRS class and the continuous score W, estimated by the system for each performance, are displayed in the right area of the window.

The second child window (Figure 6) provides a GUI that is intended to analyze and compare the performances, for each motor task, of the left and right hand but in different acquisition sessions. This GUI aims to monitor the evolution over time of the kinematic parameters that characterize the motor performance.

Figure 4. The main window GUI for the clinician subsystem.

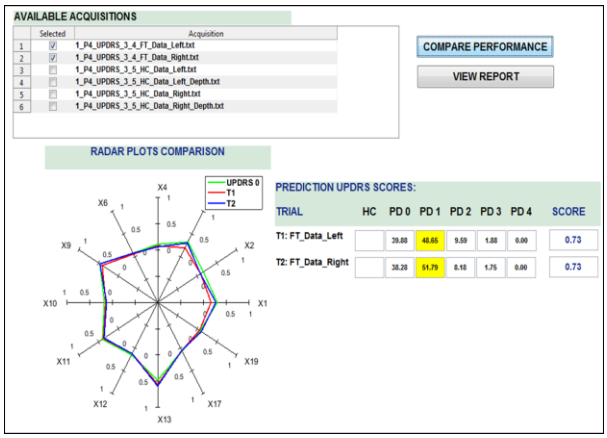


Figure 5. GUI used to compare performance of the left and right hand and to detect asymmetries. The automatic and continuous scores are displayed for each performance.

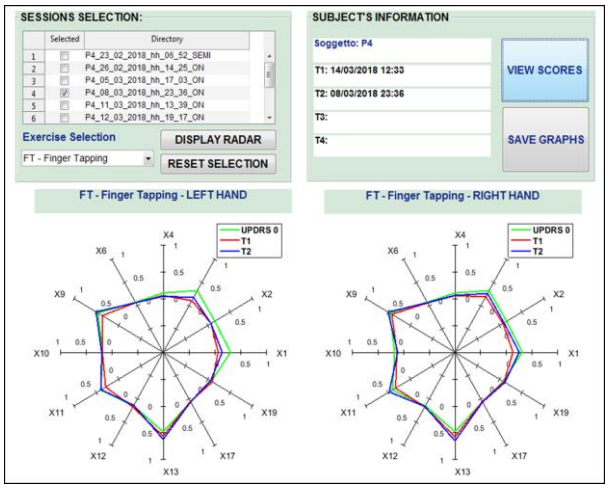


Figure 6. GUI used to compare and highlight the evolution of the performance for the left and right hand over time.

Up to four sessions can be displayed simultaneously; the related parameters can be compared immediately each other and respect to the reference parameters relating to the UPDRS 0 class. The "VIEW SCORES" button, in the upper area of the GUI, opens a third child window GUI (Figure 7). This window displays the prediction of the UPDRS classes (i.e., the output probabilities estimated by the supervised classifiers) and the continuous score W computed by the system for each performance, allowing for an easy comparison of the evolution of the patient impairment.

Finally, from the menu bar of the main window GUI, a fourth child window can be opened (Figure 8). This window allows to monitor the evolutionary trend of each kinematic parameter over time. The information displayed here may be useful to detect specific motor patterns hidden in similar performance scores, highlighting any changes in behavior over time and for both hands.

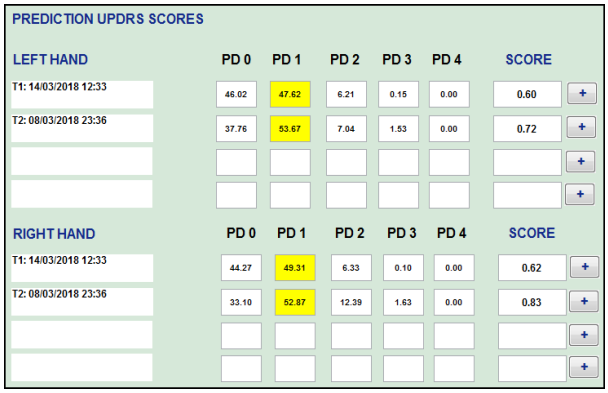


Figure 7. GUI used to display the automatic prediction and scores (output of the SVM classifier) for the selected trials, to quantify the evolution of the motor performance over time.

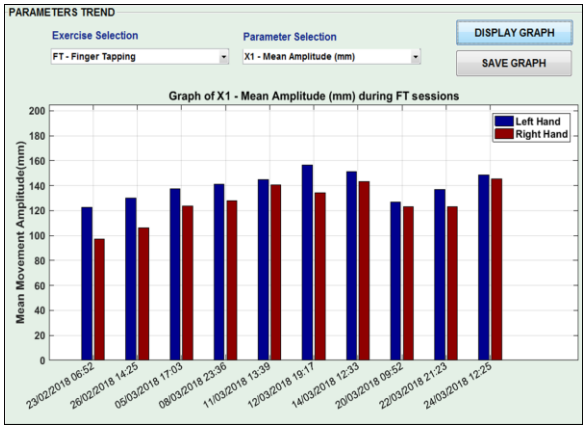


Figure 8. GUI used to display the trend of a single kinematic parameter, highlighting any change in behavior over time and for both hands.

D. Clinical Assessment and Data Acquisition

An experimental campaign was carried out to collect the kinematic data and the neurologist scores on the performances of a group of PD patients while performing the upper limb UPDRS tasks, that is Finger Tapping (FT), Opening-Closing (OC) and Pronation-Supination (PS). The goal was both to select the kinematic parameters that best characterize the differences in the impairment severity, and to collect a database of "kinematic parameters vector – neurologist UPDRS score" pairs to train the supervised classifiers of the system used for the automated assessment of each task. Two cohorts were recruited: one composed of forty patients (22 females, 18 males) with a diagnosis of Parkinson's Disease (PD), and the other composed of fifteen Healthy Control (HC) subjects. Patients were recruited according the UK Parkinson's Disease Society Brain Bank Clinical Diagnostic standards and met the following criteria: Hoehn and Yahr score (average 2.2, min 1, max 4); age 43–81 years; disease duration 2–29 years.

PD subjects were excluded if they had previous neurosurgical procedures, tremor severity > 1 (UPDRS-III severity score), or cognitive impairment (Mini–Mental State Examination Score < 27/30). The HC subjects met these criteria: age 35–78 years; not affected by neurological, motor and cognitive disorders. All subjects provided their informed consent prior to their participation.

The PD cohort was assessed for the FT, OC and PS UPDRS tasks on both hands by a neurologist experienced in movement disorders and the resulting UPDRS severity scores were found between 0 (normal) and 3 (moderate impaired). Every performance of the PD patients was tracked by the system and the related kinematic parameters were automatically extracted from the hand/fingers trajectories. The HC subjects performed the tests under the same environmental conditions and with the same system configuration as the PD patients. Before starting the acquisition campaign, a meeting was conducted to train the neurologist, staff and PD participants in the use of the system and to get acquainted with the procedures to be followed during the data acquisition.

E. Validation of the agreement between neurologist and system assessments

The goal of this work is the development of a telemedicine approach for the home-based assessment of Parkinson's Disease.

In this context, it is important to verify the agreement between the neurologist and the system assessments, both during the acquisition of experimental data and during the remote supervision, when videos of the patient performance are supervised and eventually assessed by neurologists.

The agreement between system and neurologist has been addressed using the Intra Class Correlation (ICC) coefficient [21]. The $ICC_{N\text{-}SY}$ coefficient, between the live scores assigned to each task by the neurologist and the system at the end of the patient's performance, was evaluated by applying the two-way random effects model for absolute agreement.

In addition, to verify if the video of the performance conveys enough clinical information for the remote supervision, the $ICC_{N\text{-}V}$ between live and video-based assessments was also evaluated, this by applying the two-way mixed effects model for absolute agreement.

F. Kinematic Parameter Selection

The automatic assessment of UPDRS tasks makes use of the well-established correlation existing between the kinematic parameters of the movements, objectively evaluated by the system, and the severity of the impairment, subjectively rated by neurologists and expressed as UPDRS scores [5].

The kinematic parameters we choose are closely related to the typical characteristics of the patient's movements that are used by neurologists to score the performance (amplitude, speed, rhythm, hesitations, and others). To compact the information associated with these parameters and to reduce their redundancy, the most discriminant ones

have been identified for every UPDRS task. First, the Principal Component Analysis (PCA) has been applied to the initial set of parameters to filter out those that contribute less than 5% to represent the whole dataset. Then, the selected kinematic parameters were correlated to neurologist's UPDRS scores (Spearman's correlation coefficient ρ), keeping only those with the best correlation with neurologist's UPDRS scores, at significance level $p<0.01$.

In this context, the kinematic parameters of the HC subjects have been used to normalize the PD ones. Thanks to the better performance of the HC subjects, their average score values $\mathbf{p}_{i\,HC}$ are always better than the $\mathbf{p}_{i\,PD}$ ones, and are used to obtain the set of normalized PD parameters ($\mathbf{p}_{i\,PD\,norm} = \mathbf{p}_{i\,PD} / \mathbf{p}_{i\,HC}$).

G. Automatic UPDRS Assessment by Machine Learning

To implement the automatic assessment of the FT, OC and PS UPDRS tasks, three data sets of "kinematic parameters vector – neurologist UPDRS score" pairs were used to train three different classifiers. We use the LIBSVM library package [22] to implement three Support Vector Machine (SVM) classifiers with polynomial kernel. Their accuracy in the correct assignment of the UPDRS scores was tested by using the *leave-one-out* cross validation method. The confusion matrices were used to characterize the classification performance of each SVM classifier.

An interesting feature offered by the implementation of the SVM classifier is that, given the kinematic parameters vector as input, the classifier output is a vector **P** of probabilities **p**j that the input vector belongs to the class Cj. To test the classifiers performance and build the confusion matrices, the class Ck corresponding to the highest probability **p**k among all the probabilities in **P** is chosen as the predicted score of the system.

The probabilistic assignment **P** of the classifier output allows for an interesting extension to continuous values of the discrete UPDRS classification obtained using the most probable class. For this purpose, for each task, the probabilities **p**i to belong to specific UPDRS classes (i.e., the output of the classifier) have been combined by a weighted average. In this way, a continuous estimation W of the UPDRS score is obtained:

$$W = \sum i \cdot p_i \qquad (1)$$

$i = 0..4$; p_i = probability to belong to class C_i

The advantage of this approach is the possibility of evaluating continuous and slight variations in motor impairment that is not possible to obtain with the standard quantized UPDRS score (0-4). In practice, the classifiers estimate probabilistic assignment vectors **P** having only two significant components that correspond to contiguous classes. An application to the monitoring of small fluctuations in patient impairment by the continuous UPDRS score, estimated through W, is presented in the paragraph of preliminary experiments.

H. Clinimetric validation and usability of the patient subsystem

Clinimetric validation was considered successful if a health monitoring system is shown to be reliable, valid and sensitive to changes [23][24][25]. Reliability is defined as the degree to which the measure is error-free and, consequently, produces consistent results. We use the Intra Class Coefficient (ICC) as a measure of reliability. Validity is the degree to which an instrument measures the construct it purports to measure, and we assess the system validity by the accuracy of the automated assessments as compared to the neurologist ones (Subsection C of the Results). Sensitivity is related to the ability of the system to detect changes over time. We evaluated the sensitivity in a preliminary longitudinal experiment by monitoring a limited number of PD subjects at home over a week (Subsection D of the Results). We use the continuous estimation W of the UPDRS score as defined in Equation (1) to assess the sensitivity to small changes in motor impairment.

In addition to clinimetric validation, other important aspects that a home-based and self-administered health monitoring system should show are a good usability and acceptability. For this purpose, all the recruited PD patients were interviewed at the end of their respective sessions to evaluate their global level in computer and technological skills, their ability to wear gloves, and their satisfaction in using the system. The interview was conducted by presenting them a set of adjectives qualifications referring to the system.

Furthermore, the system usability was assessed by the standardized interview of the Post-Study System Usability Questionnaire (PSSUQ) [26]. This is a 19-items ordinal score questionnaire based on 7-point Likert scales, which addresses six components of user satisfaction with regards to the systems usability: ease of use, ease of learning, simplicity, effectiveness, information and user interface.

The users' computer and technological skills were evaluated by their dichotomous answers (yes/no) to a questionnaire of 18 items concerning the previous use of information technologies (IT), the difficulties encountered in using the system, the need for a supervisor and the comprehension of the sequence of activities to be performed during the session. The users' responses to the proposed items have been added in a final score divided into 4 IT skill levels (i.e., none, basic, intermediate, advanced). The ability to wear gloves was evaluated by the session supervisor on three levels (i.e., impossible to wear, wearable with aid, wearable without aid).

The satisfaction in the use of the system was assessed by showing to the subject the image in Figure 9 at the end of the session, asking him/her to choose the three adjectives that best qualify the experience. Once again, the answers were added over all the subjects to obtain the three most important adjectives chosen by the PD cohort. The PSSUQ makes use of a standardized set of questions and procedures to evaluate the usability of the system [26].

Figure 9. Imagine shown to the users at the end of the session, with the set of proposed adjectives to qualify the experience in using the system.

IV. RESULTS

A. Discriminant kinematic parameters

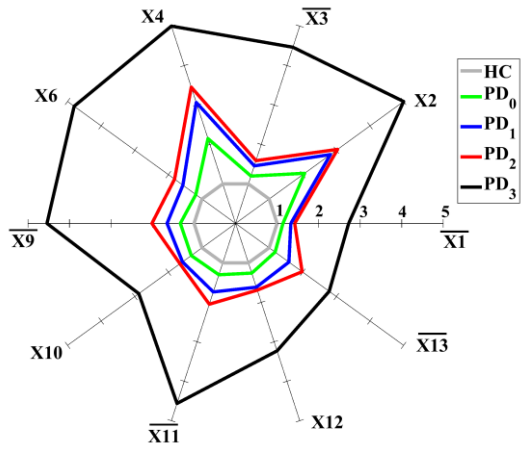
The parameter selection process, applied to the initial set of normalized parameters, produces three sets of discriminant parameters (Table I) that are able to discriminate the UPDRS classes for the FT, OC and PS tasks, respectively. This is confirmed visually by the average values of the kinematic parameters selected with respect to the UPDRS severity classes, as shown in the radar graphs of Figure 10(a) for FT, Figure 10(b) for OC and Figure 10(c) for PS tasks, respectively.

In Figure 10, the increase of the motor performance severity corresponds to an expansion of the relative radar graph. Note that, to highlight the discriminant power of the selected parameters, they have been represented directly (with the name of the original parameter) or inversely (with an overscore on the name of the original parameter), depending on whether the parameter value increases or decreases when the severity of the impairment increases. Furthermore, for graphical representation purposes, the parameters are scaled, so that the values corresponding to the best performance ($\mathbf{p}_{i\ PD\ no} = \mathbf{p}_{i\ HC}$) are represented on the innermost circle (i.e., radius value = 1).

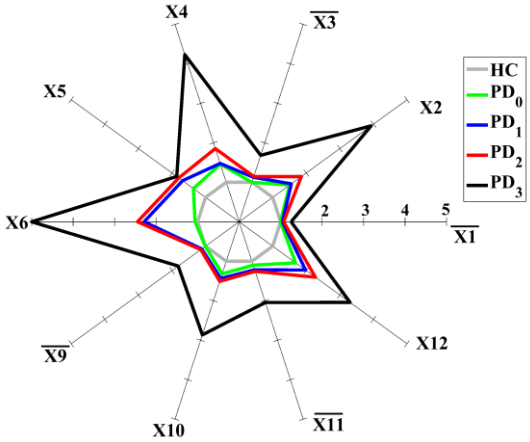
TABLE I. SELECTED KINEMATIC PARAMETERS

Name	**Finger Tapping UPDRS task**		
	Meaning	***Unit***	***ρ-value***
X_1	Maximum opening (mean)	mm	-0.43
X_2	Maximum opening (CV)	-	0.35
X_3	Maximum amplitude (mean)	mm	-0.41
X_4	Maximum amplitude (CV)	-	0.39
X_6	Duration (CV)	-	0.42
X_9	Maximum opening velocity (mean)	mm/s	-0.58
X_{10}	Maximum opening velocity (CV)	-	0.39
X_{11}	Maximum closing velocity (mean)	mm/s	-0.55
X_{12}	Maximum closing velocity (CV)	-	0.43
X_{13}	Main Frequency	Hz	-0.48
Name	**Opening-Closing UPDRS task**		
	Meaning	***Unit***	***ρ-value***
X_1	Maximum opening (mean)	mm	-0.54
X_2	Maximum opening (CV)	-	0.34
X_3	Maximum amplitude (mean)	mm	-0.55
X_4	Maximum amplitude (CV)	-	0.31
X_5	Duration (mean)	s	0.25*
X_6	Duration (CV)	-	0.58
X_9	Maximum opening velocity (mean)	mm/s	-0.63
X_{10}	Maximum opening velocity (CV)	-	0.47
X_{11}	Maximum closing velocity (mean)	mm/s	-0.54
X_{12}	Maximum closing velocity (CV)	-	0.53
Name	**Pronation-Supination UPDRS task**		
	Meaning	***Unit***	***ρ-value***
X_1	Maximum supination (mean)	deg	-0.36
X_2	Maximum supination (CV)	-	0.05
X_9	Maximum supination velocity (mean)	deg/s	-0.42
X_{10}	Maximum supination velocity (CV)	-	0.35
X_{11}	Maximum pronation velocity (mean)	deg/s	-0.46
X_{12}	Maximum pronation velocity (CV)	-	0.44
X_{13}	Main Frequency	Hz	-0.47
X_{19}	Pronation Phase Duration	s	0.33

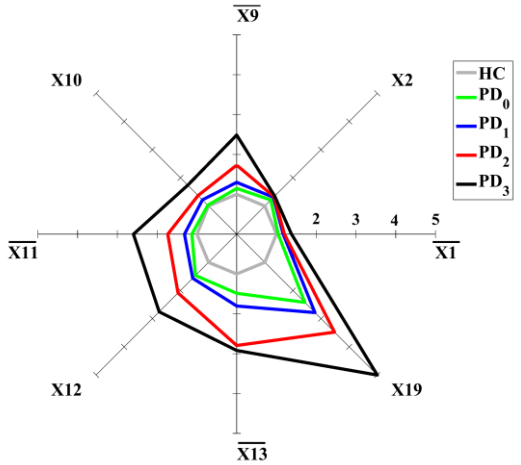
Legend
Coefficient of Variation: ratio of standard deviation (σ) to mean μ of the parameter. CV = σ/μ
Maximum Opening/Supination: peak of distance/angle in one movement
Amplitude: difference between maximum and minimum distance/angles in one movement
Duration: time elapsed between the start and the end of one movement
Maximum Opening/Supination Velocity: peak in an opening/supination phase of one movement
Maximum Closing/Pronation Velocity: peak in a closing/pronation phase of one movement
Opening/Supination Phase Duration: Time for opening/supination phase of one movement
Closing/Pronation Phase Duration: Time for closing/pronation phase of one movement
Rate: Number of movements per second
Main Frequency: Frequency with the peak in power spectrum (bandwidth 0.. 4 Hz)



(a)



(b)



(c)

Figure 10. Radar graph of selected kinematic parameters for FT task (a), OC task (b) and PS task (c)

B. *Accuracy of the Automatic Assessment*

The confusion matrices, shown in Tables II, III and IV, were used to characterize the classification performance of the SVM classifiers for the FT, OC and PS UPDRS tasks, both for the left and the right hand. From the confusion matrices, all the standard parameters for the evaluation of the classifier performance (such as accuracy, sensitivity and so on) can be easily obtained.

It can be noted that the non-zero elements outside the diagonal of the matrices are only one position far from the diagonal ones, which means that the classification errors are limited to one UPDRS class.

TABLE II. FT CONFUSION MATRIX (UPDRS CLASSES)

		SYSTEM SCORES			
		C₀	***C₁***	***C₂***	***C₃***
CLINICAL SCORES	C_0	15	3	0	0
	C_1	2	21	2	0
	C_2	0	1	18	3
	C_3	0	0	2	13

TABLE III. OC CONFUSION MATRIX (UPDRS CLASSES)

		SYSTEM SCORES			
		C₀	***C₁***	***C₂***	***C₃***
CLINICAL SCORES	C_0	14	2	0	0
	C_1	1	17	2	0
	C_2	0	1	22	3
	C_3	0	0	4	14

TABLE IV. PS CONFUSION MATRIX (UPDRS CLASSES)

		SYSTEM SCORES			
		C₀	***C₁***	***C₂***	***C₃***
CLINICAL SCORES	C_0	8	3	0	0
	C_1	1	10	2	0
	C_2	0	2	30	6
	C_3	0	0	3	15

C. *Usability assessment of the system*

The percentage breakdown of the computer skills of the PD users for none, basic, intermediate and advanced levels, is 55.2%, 18.0%, 16.8%, 10.0%, respectively. Most PD users (over 73%) had no or low computer skills, making the test representative of an elderly population of PD subjects.

The percentage breakdown of the ability to wear gloves for the three levels (i.e., impossible to wear, wearable with aids, wearable without aids) is 3%, 5%, 92%. This result

confirms that the system is quite user-friendly for people with motor impairment as PD subjects and gloves are not such an invasive equipment.

Concerning the adjectives chosen as representative of the experience with the system, the most voted ones are interesting (56%), helpful (50%), exciting (44%), stimulating (38%), unusual (25%), improvable (25%). Moreover, all the adjectives chosen have a positive meaning for the acceptability of the system. The word cloud of the characteristics expressed by the PD patients to describe their experience of using the system is shown in Figure 11. The word cloud (wordle) gives an intuitive indication of the relative importance of the adjectives proposed through the typical graphic style. The biggest words represent the most voted adjectives among the ones presented to the subjects during the interview at the end of the experimental session. Each subject was asked to selected three adjectives from a set of "positive" and "negative" words, giving us a direct feedback on the most and the least features voted in terms of usability, satisfaction and acceptability of the system.

Figure 12 shows the results of the PSSUQ questionnaire on the usability of the system, mediated on the PD cohort. Subjects were asked to answer 19 questions on the system by assigning a Likert score (1 absolute agreement, 7 absolute disagreement) to express their standard positive or negative judgement on the experience of using the system [26]. The 19 items are ordered from the first question (on the left) to the last one (on the right); for each question, the average score is reported. The analysis shows that the PD participants rated the usability of the system with an overall average score of 2.16 (±0.58) on the PSSUQ. This indicates that the majority of participants liked to use the system and appreciated the possibility offered by the system to monitor their own health condition at home.

Figure 11. The word cloud of the most voted characteristics indicated by the cohort of PD patients to describe their experience in the use of the system.

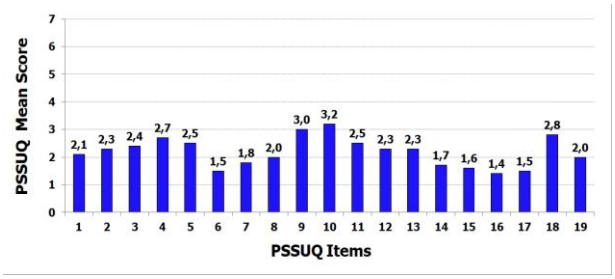


Figure 12. Results of the PSSUQ questionnaire on the usability of the system.

D. *Preliminary Experiments on UPDRS Assessment*

A preliminary experiment was conducted to assess the feasibility of the proposed system in the monitoring of PD patients at home. A small group of patients with PD (4 subjects) used the system at home for a period of one week. Subjects were instructed to perform the FT, OC and PS tasks at home every day of the week, at different times from drug intake (30m, 1.5h, 2.5h, 3.5h). The intent was to evaluate the potential fluctuations in the motor performance of upper limbs in the period following the drug intake.

Thanks to the data storage and the remote retrieving capability of the system, the test session data such as scores, parameters and, in particular, videos captured during the task execution, were remotely accessed, analyzed and evaluated by the neurologist for both the hands.

In this experiment, the agreement between the automatic scores of the system and the video-based scores of the neurologist has been evaluated by the ICC coefficient (two-way random effects model for absolute agreement). The ICC values have been evaluated for each task, collecting the four daily scores for the entire week, this for all patients. In Table V, the resulting ICC coefficients for the tasks are shown.

To give insight of the experiment results, in Figures 13, 14, and 15 are shown samples of the daily assessments by the system and by the neurologist for the FT, OC and PS tasks on the performance of a PD patient. This patient is a 65-year-old male, diagnosed for PD at 60, non-fluctuating, and with more severe motor impairment on the right side.

TABLE V. INTRA CLASS CORRELATIONS FOR THE AGREEMENT BETWEEN NEUROLOGIST AND SYSTEM SCORES

	UPDRS task		
	FT	**OC**	**PS**
ICC_{NV-SY}	0.80	0.61	0.58

The ICC values of the neurologist-system agreement for the three UPDRS tasks. The neurologist assessments are based on the recorded videos of the performances of the four patients.

The neurologist's scores are based on the recorded videos of the patient's performances: scores are evaluated and shown at four different times per day. To facilitate the interpretation, system scores expressed as continuous values W are connected by solid colored lines (red for the right hand and blue for the left hand, respectively).

In Figure 14, the large difference between left and right hand scores for the OC task, occurring at 2.5 hours from drug intake, could be due to the subjective evaluation of the neurologist. This hypothesis is supported by the other performance scores of the neurologist for the FT and PS tasks at the same time, which show less differences between the two hands. The system scores are less fluctuating, compensating for possible incorrect subjective evaluations.

As shown in the figures, on the average, there is a good agreement between system and neurologist scores. Nevertheless, the system can assess tasks on a continuous scale (W) respect to the standard discrete UPDRS score evaluated by neurologists. This feature could open the possibility to investigate the interaction between drugs and motor effects with a more objective, sensible and hopefully accurate approach.

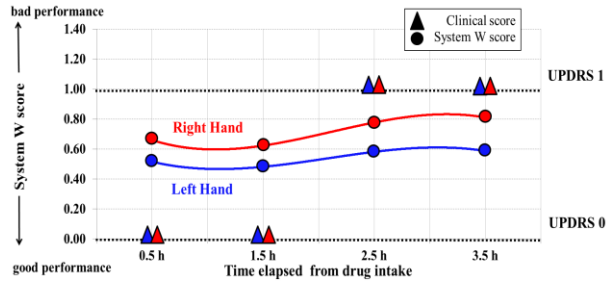


Figure 13. Example of the automatic assessment of the FT task for the left (blue) and right (red) hand at different times from drug intake. The continuous assessment scores (System W score) by the system and the standard UPDRS scores (Clinical score) by neurologist are shown at four different time. To facilitate the interpretation, system scores are also connected by solid lines.

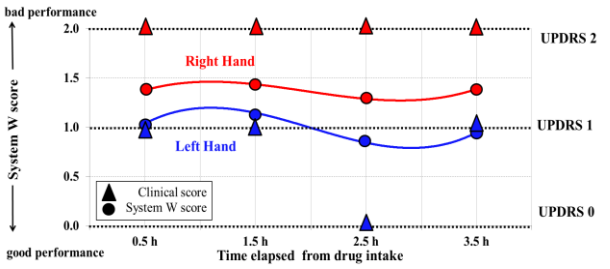


Figure 14. Example of the automatic assessment of the OC task for the left (blue) and right (red) hand at different times from drug intake. The continuous assessment scores (System W score) by the system and the standard UPDRS scores (Clinical score) by neurologist are shown at four different time. To facilitate the interpretation, system scores are also connected by solid lines.

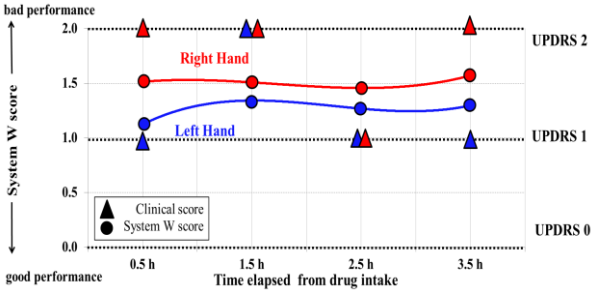


Figure 15. Example of the automatic assessment of the PS task for the left (blue) and right (red) hand at different times from drug intake. The continuous assessment scores (System W score) by the system and the standard UPDRS scores (Clinical score) by neurologist are shown at four different time. To facilitate the interpretation, system scores are also connected by solid lines.

V. CONCLUSIONS AND FUTURE WORKS

This work presents a non-invasive and low-cost system for the automatic assessment of subjects with PD that perform standard UPDRS tasks for upper limbs at home. The system is based on a new human computer interface that, through the accurate hand tracking, allows both the management of the system and the automatic and objective UPDRS assessment.

The gestural interface makes it suitable for users with motor impairment, as are PD patients. The user interface of the system has been specifically designed for the home monitoring of people with mobility difficulties, as those affected by Parkinson's Disease.

The system interface of the remote supervisor provides a secure access to the clinical data. Furthermore, all relevant clinical data (videos, reports with automatic scores and information entered by the patient) are displayed and can be easily analyzed by the clinician. Textual messaging can be used by the remote supervisor to send messages, which are shown on the GUI of the patient subsystem at the start of the next acquisition session.

The automatic assessment of UPDRS tasks is performed by a machine learning approach that uses some selected kinematic parameters that characterize the patient's movements. The classifiers, one for each UPDRS task, were trained during an experimental campaign in which patients with PD were assessed at the same time by the neurologist and the system. The results obtained from the classifiers confusion matrices show that classification errors are limited to one UPDRS class and only in some cases, making the system suitable for the self-managed assessment of the upper limbs UPDRS tasks at home. Based on the classifier output, a new continuous estimate of the UPDRS score is introduced and its potential benefit is discussed.

The clinimetric properties of the system and its usability for PD users have been evaluated. The results confirm that the system is suitable for the monitoring of Parkinson's Disease at-home.

Preliminary results on the application of the continuous UPDRS score in the at home monitoring of patients with PD are presented. Further experiments are still needed to validate both the usability and accuracy of the system in home environment, and the usefulness of the continuous UPDRS score introduced here for monitoring fine fluctuations of motor impairment.

Next steps will also cover the extension of this solution to the analysis of other UPDRS tasks, with the aim of obtaining a complete and comprehensive assessment of the neuro-motor status of PD patients. This would be important in the perspective of the optimization of the drug therapy, because the assessments could be carried out on demand at the patient's home whenever more frequent observations are needed to assess the worsening of motor symptoms. All these features are relevant to significantly improve both the clinical management and the patient's quality of life.

REFERENCES

[1] R. Nerino, C. Ferraris, G. Pettiti, A. Chimienti, C. Azzaro, G. Albani, L. Priano, and A. Mauro, “Steps Toward Automatic Assessment of Parkinson’s Disease at Home”, 4th International Conference on Smart Portable, Wearable, Implantable and Disability-oriented Devices and Systems, SPWID 2018, Barcelona, Spain, July 2018 , pp. 15-20

[2] G. Pal and C. G. Goetz, “Assessing bradykinesia in Parkinsonian disorders”, Front. Neurol., vol. 4(54), Jun. 2013, pp. 1-5, eCollection, doi:10.3389/fneur.2013.00054

[3] Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease, “The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations”, Mov. Disord., vol. 18(7), Jul. 2003, pp. 738-50, doi:10.1002/mds.10473

[4] M. Richards, K. Marder, L. Cote, and R. Mayeux, "Interrater reliability of the Unified Parkinson's Disease Rating Scale motor examination", Mov. Disord., vol. 9(1), Jan. 1994, pp. 89-91

[5] D. A. Heldman, J. P. Giuffrida, R. Chen., M. Payne, F. Mazzella, A. P. Duker, A. Sahay, S. J. Kim, F. J. Revilla, and A. J. Espay, “The Modified Bradykinesia Rating Scale for Parkinson’s Disease: Reliability and Comparison with Kinematic Measures”, Mov. Disord., vol. 26(10), Aug. 2011, pp. 1859-1863, doi:10.1002/mds.23740

[6] A. L. Taylor Tavares, G. S. Jefferis, M. Koop, B. C. Hill, T. Hastie, G. Heit, and H. M. Bronte-Stewart, “Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation”, Mov. Disord., vol. 20(10), Oct. 2005, pp. 1286-1298, doi:10.1002/mds.20556

[7] A. J. Espay, P. Bonato, F. B. Nahab, W. Maetzler, J. M. Dean, J. Kluken, and B. M. Eskofier Movement Disorder Society Task Force on Technology, "Technology in Parkinson's disease: Challenges and opportunities", Mov. Disord., vol. 31(9), Sep. 2016, pp. 1272-1282, doi:10.1002/mds.26642

[8] S. Patel, K. Lorincz, R. Hughes, N. Huggins, J. Growdon, D. Standaert, M. Akay, J. Dy, M. Welsh, and P. Bonato, “Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors”, IEEE Trans Inf Technol Biomed., vol. 13(6), Nov. 2009, pp. 864-873, doi:10.1109/TITB.2009.2033471

[9] A. J. Espay, J. P. Giuffrida, R. Chen, M. Payne, F. Mazzella, E. Dunn, J. E. Vaughan, A. P. Duker, A. Sahay, S. J. Kim, F. J. Revilla, and D. A. Heldman, “Differential Response of Speed, Amplitude and Rhythm to Dopaminergic Medications in Parkinson’s Disease”, Mov. Disord., vol. 26(14), Dec. 2011, pp. 2504-2508, doi:10.1002/mds.23893

[10] J. Stamatakis, J. Ambroise, J. Cremers, H. Sharei, V. Delvaux, B. Macq, and G. Garraux, “Finger Tapping Clinemetric Score Prediction in Parkinson’s Disease Using Low-Cost Accelerometers”, Comput Intell Neurosci, vol. 2013, 2013, pp. 1-13, doi:10.1155/2013/717853

[11] T. O. Mera, D.A. Heldman, A. J. Espay, M. Payne, and J. P. Giuffrida, "Feasibility of home-based automated Parkinson's disease motor assessment”, J Neurosci Methods, vol. 203(1), Jan. 2012, pp. 152-156, doi:10.1016/j.jneumeth.2011.09.019

[12] N. P. Oess, J. Wanek, and A. Curt, “Design and evaluation of a low-cost instrumented glove for hand function assessment”, J Neuroeng Rehabil, vol. 9(2), Jan. 2012, doi:10.1186/1743-0003-9-2

[13] T. Khana, D. Nyholmc, J. Westina, and M. Dougherty, “A computer vision framework for finger-tapping evaluation in Parkinson's disease”, Artif Intell Med, vol. 60(1), Jan. 2014, pp. 27-40, doi:10.1016/j.artmed.2013.11.004

[14] A. Jobbagy, P. Harcos, R. Karoly, and G. Fazekas, “Analysis of finger-tapping movement”, J Neurosci Meth, vol. 141(1), Jan. 2005, pp. 29-39, doi:10.1016/j.jneumeth.2004.05.009

[15] C. Ferraris, R. Nerino, A. Chimienti, G. Pettiti, D. Pianu, G. Albani, C. Azzaro, L. Contin, V. Cimolin, and A. Mauro, “Remote monitoring and rehabilitation for patients with neurological diseases”, Proc. 10th International Conference on Body Area Networks (BodyNets 14), Oct. 2014, pp. 76-82, doi:10.4108/icst.bodynets.2014.257005

[16] A. H. Butt, E. Rovini, C. Dolciotti, P. Bongioanni, G. De Petris, and F. Cavallo, “Leap motion evaluation for assessment of upper limb motor skills in Parkinson's disease”, IEEE International Conference on Rehabilitation Robotics (ICORR 17), Jul. 2017, pp. 116-121, doi: 10.1109/ICORR.2017.8009232

[17] P. J. M. Bank, J. Marinus, C. G. M. Meskers, J. H. de Groot, and J. J. van Hilten, "Optical Hand Tracking: A Novel Technique for the Assessment of Bradykinesia in Parkinson's Disease", Mov. Disord. Clinical Practice, vol. 4(6), Dec. 2017, pp. 875-883, doi:10.1002/mdc3.12536

[18] B. Dror, E. Yanai, and A. Frid, "Automatic assessment of Parkinson's Disease from natural hands movements using 3D depth sensor", IEEE 28th Convention of Electrical & Electronics Engineers in Israel (IEEEI), Dec. 2014, doi: 10.1109/EEEI.2014.7005763

[19] C. Ferraris, D. Pianu, A. Chimienti, G. Pettiti, V. Cimolin, N. Cau, and R. Nerino, "Evaluation of Finger Tapping Test Accuracy using the LeapMotion and the Intel RealSense Sensors", IEEE 37th Annual International Conference of Engineering in Medicine and Biology Society (EMBC 15), Aug. 2015, one page abstract

[20] G. Bradski and A. Kaehler, “Learning OpenCV: Computer Vision with the OpenCV Library”, O'Reilly Media, 1 ed., Oct. 2008, pp. 1-580

[21] M. Richards, K. Marder, L. Cote, and R. Mayeux, "Interrater reliability of the Unified Parkinson's Disease Rating Scale motor examination", Mov. Disord., 1994; 9(1): 89-91

[22] C. C. Chang and C. J Lin, "LIBSVM: a library for support vector machines", ACM transactions on Intelligent Systems and Technology (TIST), vol. 2(3), Apr. 2011, pp. 27:1-27:27, doi: 10.1145/1961189.1961199

[23] C. Godinho, J. Domingos, G. Cunha, A.T. Santos, R.M. Fernandes, D. Abreu, and A. Al-Jawad, “A systematic review of the characteristics and validity of monitoring technologies to assess Parkinson’s disease”, Journal of neuroengineering and rehabilitation, 2016, 13(1), 24

[24] L.B. Mokkink, C.B. Terwee, D.L. Patrick, J. Alonso, P.W. Stratford, D.L. Knol, et al. “The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes”, J Clin Epidemiol. 2010, 63, pp. 737–45, doi:10.1016/j.jclinepi.2010.02.006

[25] J.J. Ferreira, C. Godinho, A.T. Santos, J. Domingos, D. Abreu, R. Lobo, and I. Akeren, “Quantitative home-based assessment of Parkinson’s symptoms: the SENSE-PARK feasibility and usability study”, BMC neurology, 2015, 15(1), 89

[26] J.R. Lewis, "Psychometric evaluation of the PSSUQ using data from five years of usability studies", International Journal of Human-Computer Interaction, 2002, 14(3-4), pp. 463-488

The Role of Adaptive Immersive Technology in Creating Personalised Environments for Emotional Connection and Preservation of Identity in Dementia Care

Insights from User Perspectives towards SENSE-GARDEN

Gemma Goodall[1], Ileana Ciobanu[2], Ronny Broekx[3], Jon Sørgaard[1], Iulian Anghelache[4], Catalina Anghelache-Tutulan[4], Mara Diaconu[4], Sigrid Mæland[5], Therese Borve[5], Audun Digranes Dagestad[5], Piet Bormans[6], Marleen Custers[6], Katrin Losleben[7], Rita Valadas[8], Cristina Vaz de Almeida[8], Alda Matias[8], Andreea Marin[2], Kristin Taraldsen[1], Walter Maetzler[9], Mihai Berteanu[2], J. Artur Serrano[1,10]

[1]Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, NTNU/Norwegian University of Science and Technology, Norway
[2]University of Medicine and Pharmacy Carol Davila, ELIAS University Hospital, Bucharest, Romania
[3]Innovation Department, ePoint, Belgium
[4] Research & Development Department, Compexin SA, Romania
[5]Odda Municipality, Norway
[6]Aan de Beverdijk, Vulpia Vlaanderen, Belgium
[7]Center for Women's and Gender Research, UiT The Arctic University of Norway
[8]Santa Casa da Misericórdia de Lisboa, Portugal
[9]Department of Neurology, Christian-Albrechts-University, Kiel, Germany
[10]Norwegian Centre for eHealth Research, University Hospital of North Norway

Email: gemma.goodall@ntnu.no, ileanacuk@yahoo.co.uk, ronny.broekx@gmail.com, jon.sorgaard@ntnu.no, a.iulian@compexin.ro, catalina.anghelache@compexin.ro, mara.diaconu@compexin.ro, sigrid.meland@odda.kommune.no, therese.borve@odda.kommune.no, audun.digranes.dagestad@odda.kommune.no, piet.bormans@beverdijk.be, marleen.custers@beverdijk.be, katrin.losleben@uit.no, rita.valadas@scml.pt, cristina.vazalmeida@scml.pt, alda.matias@scml.pt, andreea.budrica@gmail.com, kristin.taraldsen@ntnu.no, w.maetzler@neurologie.uni-kiel.de, mberteanu@gmail.com, artur.serrano@ntnu.no

Abstract— **This paper presents early stage research on the development of an immersive, multisensory room for people living with dementia. Dementia is considered to be a public health priority on a global level. Our research addresses the challenge of meeting individual needs in dementia care, particularly in relation to social and emotional wellbeing. We draw upon findings from 52 interviews with users, including people with mild cognitive impairment, professional caregivers, and informal caregivers. These interviews were conducted to explore initial responses towards a personalised multisensory room called SENSE-GARDEN. Thematic analysis resulted in six themes: benefits for all, focus on the individual, past and present, emotional stimulation, shared experiences, and challenges to consider. This paper provides important theoretical considerations for the role of technology in not only the SENSE-GARDEN intervention, but in preserving the identities of people with dementia and providing opportunities for connection with others. Future work in this area should adopt an interdisciplinary approach to using technology in dementia care.**

Keywords-dementia; virtual environments; immersive technology; human computer interaction; interpersonal relationships

I. INTRODUCTION

This article builds upon a conference paper presented at the Fourth International Conference on Human and Social Analytics [1]. This extended version of the original paper offers detailed results from a preliminary study on a virtual adaptive environment for people with dementia.

Dementia is a syndromal term and can be caused by a variety of diseases, including neurodegenerative diseases. Memory, behaviour, and communicative abilities are often affected [2]. There are approximately 47 million people living with dementia worldwide [3]. With this number set to increase to 131.5 million by 2050, it is of the utmost importance to tackle dementia's progressive impact on the wellbeing of people living with this syndrome.

The World Health Organization has called for action on dementia, presenting it as a public health priority at a global level [2]. This action includes a call for research to identify ways of supporting the needs of people living with dementia, their caregivers, and the needs of society in the context of costs, understanding, and awareness.

In recent years, studies have identified numerous complex needs of people with dementia living in long-term care. These include the management of challenging behaviours, maintenance of social relationships, involvement

of people with cognitive deficits in meaningful activities, and supporting the emotional needs of all [4][5].

Emotion-oriented approaches to care have been shown to be cost-effective ways of improving psychological wellbeing and social behaviour amongst people with dementia [6][7]. These nonpharmacological approaches are often person-centred, focusing on the social and emotional needs of the individual. Reminiscence rooms, virtual gardens and virtual reality forests are examples of how immersive technologies have been integrated in emotion-oriented approaches designed to create effective nonpharmacological interventions for people with dementia [8][9].

However, this area of study has called for further research in determining what works best for the individual [10]. It has recently been suggested that an individualised multisensory environment for people with dementia would be a highly beneficial intervention, especially if family members are included in the selection of stimuli [11]. Our research is in line with this suggestion, creating not only a personalised multisensory space and intervention, but one that also incorporates immersive technology, all with the inclusion of family members, friends, and professional care staff.

This paper presents early stage research on a multisensory room, SENSE-GARDEN, that is currently being developed as an adaptive, immersive environment integrating technology and multisensory stimulation for reminiscence in people living with mild to moderately severe dementia. We will first provide a brief overview of the project (Section II), followed by a description of the methodology used in research and development (Section III). We will then discuss the results of the interviews in relation to each of the six themes identified through thematic analysis (Section IV). In Section V, the results are summarised and discussed in relation to the role of technology in preserving the identity of the person with dementia and facilitating an environment in which relationships can be fostered. Finally, in Section VI, we conclude with final remarks, the next steps for SENSE-GARDEN, and suggestions for future research.

II. SENSE-GARDEN: An Overview

SENSE-GARDEN is a psychosocial intervention that is being developed to create individualised reminiscence sessions for people living with dementia in residential care. The intervention combines the use of technology for reminiscence and multisensory stimulation, with human-to-human informational and emotional communication.

Prototypes of the SENSE-GARDEN room are currently being built across several countries in Europe, namely in Norway, Portugal and Romania, with an initial prototype already being tested in Belgium. These rooms are filled with individualised stimuli such as familiar music, soundscapes, imagery, films, and scents in order to stimulate memory and encourage active participation of the person with dementia in reminiscing activities. Particular emphasis is placed on using autobiographical content such as family photographs, music from childhood, and films of life events.

The use of large projection screens, scent dispensers, and surround sound systems will integrate the various multimedia of the room, creating an immersive environment. For example, high-definition imagery of a forest could be accompanied with the smell of pine trees and the sound of birds, to evoke a completely immersive sensation.

SENSE-GARDEN will expand on currently established sensory rooms, which are also known as ‘Snoezelen’ rooms. Deriving from the Dutch terms for ‘sniffing’ and ‘dozing’, Snoezelen was originally developed in the Netherlands as a therapy for individuals with learning difficulties [12].

SENSE-GARDEN presents an innovative approach to sensory rooms by utilising smart technologies that enable the space to adapt to the individual preferences and needs of the person with dementia. This focus on autobiographical content is achieved through the use of individual user profiles. Each profile has an associated media repository consisting of digital photographs, films, and music that holds significant meaning for the person with dementia.

Radio frequency identification (RFID) is used to allow the SENSE-GARDEN system to identify the user. Upon entering the room, the system automatically projects autobiographical multimedia from the person with dementia’s user profile.

The room is designed to be used by two main categories of users. The first is the person with dementia (PwD), who is also considered the primary user. The second is the caregiver, who will either be informal (family/friend) or formal (professional care staff). It is anticipated that together, the PwD-caregiver dyad will interact with the immersive environment to stimulate memory, conversation, sharing and engagement.

III. Methodological Approach

SENSE-GARDEN is a multidisciplinary project involving partners in Belgium, Norway, Portugal, and Romania. The consortium brings together multiple professions and competencies including technology development, architecture, care home management, health sciences and research.

There have been numerous calls to involve people with dementia in the process of designing assistive technologies [13][14]. Their contributions are thought to be of crucial importance, along with input from their caregivers [15]. More recently, user centred design has been recommended for the development and implementation of psychosocial interventions [16].

The SENSE-GARDEN project embraces a user centred design approach and is working co-creatively with user groups throughout all its phases. The aim of this preliminary research was to explore initial responses from user groups, so that their ideas and feedback may be integrated into the next phases of development of SENSE-GARDEN.

Thus far, 52 qualitative semi-structured interviews have been conducted with user groups across Belgium, Norway, Portugal, and Romania. The aims of these interviews were to collect responses and attitudes towards the SENSE-GARDEN room concept, and to identify challenges that may arise during the course of the project.

TABLE I. RESPONDENT INFORMATION

Country	People with Mild Cognitive Impairment				Informal Caregivers				Formal Caregivers			
	N	Mean Age	Gender		N	Mean Age	Gender		N	Mean Age	Gender	
			Male	Female			Male	Female			Male	Female
Belgium	3	89.6	2	1	6	57	1	5	4	31.5	1	3
Norway	4	84	0	4	4	59.3	0	4	4	38.8	1	3
Portugal	3	79.7	0	3	3	55.7	0	3	3	44.3	0	3
Romania	6	67.2	3	3	6	50.7	0	6	6	42.7	2	4
Total	**16**	**77.9**	**5**	**11**	**19**	**55.3**	**1**	**18**	**17**	**39.4**	**4**	**13**

The specific research questions for this study were as follows: (1) What are the users' attitudes towards the concept of SENSE-GARDEN? (2) What benefits, if any, do users think SENSE-GARDEN could provide in the care of people living with dementia?

In order to answer these research questions, the interview was designed in a way that allowed for an in-depth exploration of the users' beliefs surrounding SENSE-GARDEN. The interview was semi-structured with open-ended questions and lasted for approximately 30 minutes. Interview questions focused on the overall concept of SENSE-GARDEN, the individual components of the intervention, and potential benefits.

The respondents included 16 people living with a diagnosis of mild cognitive impairment, 19 informal caregivers, and 17 professional caregivers. Table 1 gives an overview of the respondent information.

In order to conduct an in-depth exploration of the ideas and perspectives given by the users, data was analysed using thematic analysis. Thematic analysis is a qualitative method in which prevalent patterns of ideas and responses are identified amongst data. The analysis procedure for this study undertook the following phases, given by Braun and Clarke [17]:

1) Familiarisation with the data: All the data was thoroughly read and re-read, along with notating initial ideas and interpretations of the dataset.

2) Coding: The ideas were used to generate codes, which identify interesting features across the data. In this study, data was manually coded in an inductive manner, meaning that the codes and themes were developed directly from the content of the data, rather than being developed by pre-existing ideas.

3) Searching for themes: The codes were used to search for themes, which represent patterned responses or meanings across the data.

4) Reviewing themes: The themes were reviewed to ensure that they accurately represent the views of the users and the view from the entire dataset.

5) Defining and naming themes: The essence of each theme was identified, along with its relevance to the research questions.

6) Producing the report: Finally, the themes were considered in their relationship to one another, and a narrative about the dataset was created. This narrative is supported by direct quotes from the dataset.

In order to stay true to the 'voice' of the users, codes and themes were constantly checked back against original data. Braun and Clarke [17] emphasise the importance of flexibility in thematic analysis and identify the process as one of continuous reflection on the reading, shaping, and checking of data and themes.

IV. RESULTS

Six themes were identified through the thematic analysis: *(A)* Benefits for All, *(B)* Focus on the Individual *(C)* Past and Present, *(D)* Emotional Stimulation, *(E)* Shared Experiences, and *(F)* Challenges to Consider. A thematic map is shown in Figure 1 to provide a visual summary of all six themes and their respective subthemes.

This thematic map also demonstrates the interactive nature of the themes and their relationship to one another. Numerous subthemes falling under different main themes are related to each other. For example, the subtheme of 'stimulating emotional memory' (under the theme of Emotional Stimulation) can be connected to the subtheme of 'avoiding negative memories' (under the theme of challenges to consider). In this way, all the themes presented provide an overarching narrative of the users' beliefs, views, and attitudes towards SENSE-GARDEN and the technology within it.

The following subsections will discuss each of the six themes in turn. The full dataset from the interviews has been made available online, along with the interview guide, and coding from thematic analysis [18].

A. Benefits for All

There was a resounding view from all users that SENSE-GARDEN may be able to provide benefit in some way. These benefits were grouped into five subthemes: benefits for the person with dementia, benefits for the family, benefits for professional caregivers, benefits in practice, and benefits beyond dementia care.

Benefits for the person with dementia. All users believed that SENSE-GARDEN has the potential to provide numerous benefits for people living with dementia. These benefits included improvements in memory, mood and overall quality of life: "Stimulating memory and improving quality of life, the person with dementia and caregivers can enjoy life more"..., "This can enrich their [people with dementia] everyday life"...,"I am sure this will be of value.

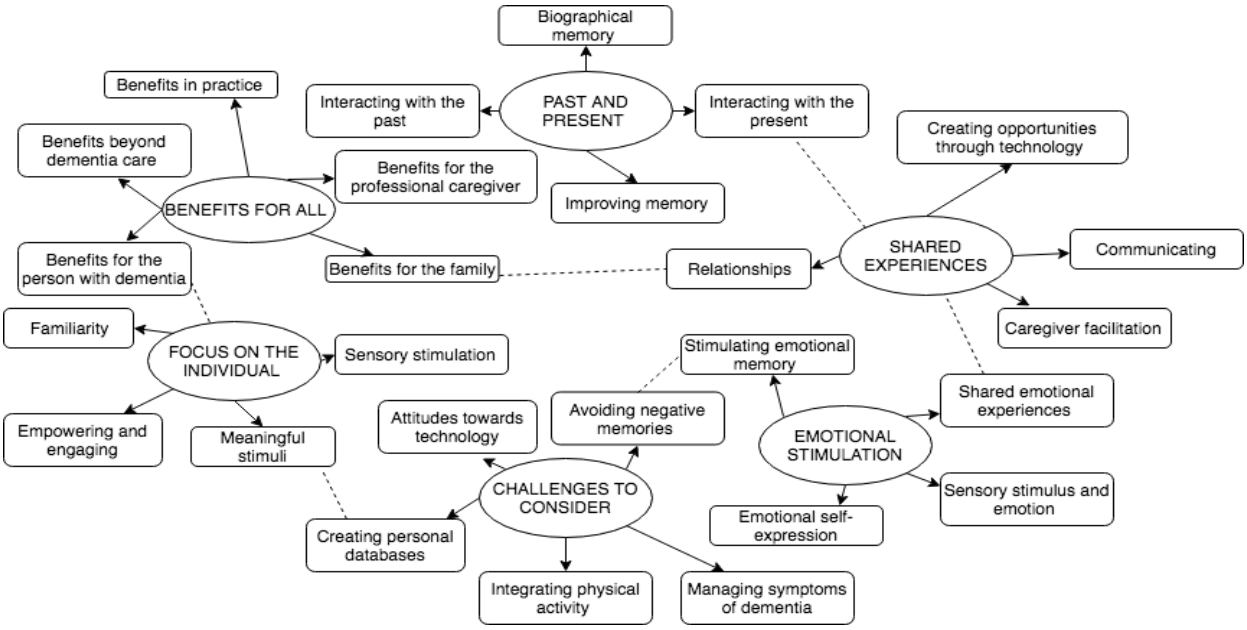


Figure 1. Thematic map of themes and subthemes identified across the dataset.

The person with dementia gets a good experience every day. In this we have faith". One person with mild cognitive impairment discussed the role of the intervention in tackling issues of helplessness that are associated with not only dementia, but illness in general: "When ill, it is like you are closed in a dark place you cannot leave by yourself. SENSE-GARDEN can help you out".

Benefits for the family. Many of the caregivers, both informal and formal, commented on SENSE-GARDEN being able to provide ways for the family to strengthen relationships with loved ones who have dementia: "It's hard to be a relative, so little competence, dialogue is difficult. This [SENSE-GARDEN] is a great tool for having a nice time together".

Benefits for the professional. Formal caregivers considered SENSE-GARDEN as a tool for getting to know people with dementia better. The highly personalised nature of the intervention means that staff have the opportunity to gain insight into the resident's life in a way that is perhaps not possible in day-to-day care: "The advantage is that you can have full focus on the patient, being able to be alone with him or her. We get to know the patient better. It creates security."..., "This will also mean that the staff become better acquainted with the person with dementia". Another caregiver commented as follows: "It's good for the staff to see the person with dementia in another way". These comments go to suggest that digital media can create opportunities for learning more about individuals with dementia, which could be especially important for people in later stages of dementia, who may not be able to coherently express themselves.

Benefits in practice. As well as presenting individual benefits, users believed that SENSE-GARDEN could benefit the healthcare system in terms of cost and practice: "Why has nobody thought of this before? Many of these things should have already been at the nursing home even if one does not have a SENSE-GARDEN"..., "May become important in terms of reducing the cost of dementia care over time".

Benefits beyond dementia care. There was a consensus across the respondents that SENSE-GARDEN could also provide benefits to people living without dementia: "It is always good to go back to childhood and youth, for all of us. No need to be a person with dementia".

The users' positive outlook on SENSE-GARDEN captures a range of benefits that not only apply to the person with dementia, but also to caregivers and care practice as a whole. Future studies on SENSE-GARDEN will need to incorporate outcome measures that evaluate these various aspects.

B. *Focus on the Individual*

The key concept of SENSE-GARDEN is creating an environment in which the person with dementia is the central focus. The users not only valued this focus on the person with dementia, but they also offered their suggestions on to how best create an individualised environment. These suggestions are grouped into the following subthemes: familiarity, meaningful stimuli, sensory stimulation, and empowering and engaging.

Familiarity. With the SENSE-GARDEN being a new and unfamiliar concept, both informal and formal caregivers stressed the importance of providing a familiar surrounding for the person with dementia: "A familiar environment, familiar objects to touch, is mandatory"..., "At least for the first sessions, the SENSE-GARDEN room must include

familiar items, besides the personal records used for projection and music".

Meaningful stimuli. Users believed that the stimuli used in SENSE-GARDEN should have significant meaning for the person with dementia: "Family photo album, with photos from important emotional occasions"..., "Meaning from one's own trips. You must remember a trip, but also the reason you went on that trip, the scope".

Sensory stimulation. There was an overall positive attitude towards SENSE-GARDEN's proposed methods of sensory stimulation. Users commented on the ability for such stimulation to trigger memories and improve mood: "Imagine what scent can bring forth, the idea of what this can do, it's gorgeous". There were also numerous suggestions for SENSE-GARDEN to broaden its current plans for sensory stimuli, such as including tactile elements: "Maybe something more for the sense of touch. When you see a mountain and smell the fern tree, why not touch a fern tree branch?".

Empowering and engaging. SENSE-GARDEN was perceived as an opportunity for people with dementia to actively engage and express themselves: "The person with dementia has to be reassured that life has not come to an end when diagnosed with dementia, and reality is not limited by the walls of the bedroom. They still have things to show and share with us all".

The suggestions given by the users imply that whilst the technology and media within the SENSE-GARDEN needs to be individualized, there are additional ways in which individualisation can be achieved. This is through caregiver facilitation, tactile stimuli, and the physical design of the room. All of these factors will need to be taken into consideration throughout the development of the intervention.

C. Past and Present

Given that SENSE-GARDEN borrows techniques from reminiscence therapy, it is of no surprise that discussion regarding memories arose during the interviews. However, the users identified links between interaction with the past and with the present, as well as the impression of overall improvement of memory in general. Therefore, the subthemes are: interacting with the past, interacting with the present, and improving memory over sessions.

Interacting with the past. In discussing the benefits of SENSE-GARDEN, all respondents believed that the individualised nature of the virtual environment could trigger autobiographical memories. This was linked to helping people with dementia connect with their past: "Personal videos and photos are important. You resonate with your past".

Improving memories over sessions. As well as stimulating memories of the past, respondents also believed that memory could be strengthened over the course of the SENSE-GARDEN sessions. Some users suggested using visual markers in the SENSE-GARDEN components in order to trigger memory in consequent sessions: "Using memory anchors will improve experience and stimulate reality connection". An example of this would be to use a recent photograph of a familiar place that holds significant meaning for the person with dementia. The same photograph could then be presented to the user in the next SENSE-GARDEN session to see if they remember the meaning connected to that picture.

Interacting with the present. There was a suggestion that even if the person with dementia does not have the capacity for long term memory of the sessions, the individual could still benefit from the 'in-the-moment' experience of SENSE-GARDEN: "They probably do not remember afterwards, but think about being happy one hour every day. That's a good benefit". Respondents also considered interaction with the past an activity for strengthening self-identity in the present moment: "Nowadays we forget who we are. SENSE-GARDEN will help us all relive forgotten events and identities".

This symbiotic relationship between past and present has been much discussed in regards to selfhood. Surr [19] adopts a socio-biographical approach to explain how people with dementia use their past in the context of telling their life story to others, in order to maintain a sense of self in the present. Technology may have much to offer in this maintenance of self, ideas of which will be given in detail in the discussion section of this paper.

D. Emotional Stimulation

Whilst emotion was a prominent topic amongst all of the themes, the comments from the users proved emotion to be highly complex. It was therefore decided to include a more detailed discussion of emotion. The subthemes are as follows: sensory stimulus and emotion, stimulating emotional memory, emotional self-expression, and shared emotional experiences.

Sensory stimulus and emotion. The users believed that stimulating the senses through imagery and music could stimulate positive emotions in the person with dementia: "One connects so much to music, there are a lot of emotions"..., "Stimulating senses brings joy and memories".

Stimulating emotional memory. The users focused primarily on familiar music in being able to stimulate emotional memory in the person with dementia. "Just three notes will bring back that special moment if music is connected to that moment"....,"When we hear a song, we think of something and then we will be happy".

Emotional self-expression. Individuals with dementia are capable of experiencing and expressing a wide range of emotions, even in later stages of the disease [6][20]. Building upon the idea of sensory stimulation triggering emotional memories, the users also believed that SENSE-GARDEN could enable people with dementia to express themselves in ways that transcend typical verbal communication: "Some people stop talking, but they can sing". Furthermore, they believed that people may be able to experience a heightened state of feeling through the intervention: "SENSE-GARDEN is an intermediary space, between the memories and the here and now, a space we can all access and we can remember how to feel, by one's self and together, without shame or fear".

Shared emotional experiences Finally, the discussion of emotion went beyond individual feelings. The users expressed the value of SENSE-GARDEN in being able to help people connect with one another: "Sharing the experience is most important for reconnecting". One person with mild cognitive impairment also highlighted the importance of how these shared experiences should be shaped: "The therapist is very important and can instil peace and wellbeing. A special emotional environment must be created for SENSE-GARDEN to work." The idea of creating a "special emotional environment" goes to suggest that it is not the intervention alone that can provide benefits to the relationships, but it is also the individuals present who can shape the experience of SENSE-GARDEN.

This theme has demonstrated the intricate nature of emotions, and how they can be manifested through the stimulation of the senses, through the remembrance of past events, and through our relationships with others.

E. Shared Experiences

SENSE-GARDEN is designed to be a joint experience between the person with dementia and their caregiver. As discussed in the previous subtheme, the users expressed the importance of sharing the experience together. This current theme goes beyond that of emotions and discusses the shared experience in relation to the following subthemes: caregiver facilitation, relationships, communicating, and creating opportunities through technology.

Caregiver facilitation. Many users believed that carefully planned facilitation of SENSE-GARDEN is required for the intervention to work. Particular stress was placed on the importance of being accompanied by a familiar individual: "We must have people accompany us- internal people we know". Users also believed that effective facilitation could shape a positive environment in which the benefits of the intervention could be maximised: "The caregiver must be well trained and possess good communication skills...to be able to support and fructify the person with dementia's gains in terms of cognitive and behavioural improvements".

Relationships. Respondents believed that SENSE-GARDEN could improve understanding and relationships between people with dementia and their caregivers- both formal and informal. There was a sense of the intervention being able to 'restore' what dementia had taken away from the relationship, such as self-identify and communication: "Family and friends can be with the patient as they were before". SENSE-GARDEN was considered a catalyst for fostering relationships and providing opportunities for self-expression and understanding between people with dementia and their families. This improvement in relationships was considered important in easing caregiver burden: "Improving relationships with family members and staff, easing caregiver burden on the staff and family".

Communicating. During the interviews, discussions turned to benefits of creative activity in dementia care. In particular, there were strong references to the ability of visual media and music to provide alternative forms of communication beyond that of verbal means. Users believed that the inclusion of music and visual imagery in SENSE-GARDEN would be able to provide tools for sharing information: "Being able to tell stories, if one has lost the language, pictures and movies can tell things." Users also believed that SENSE-GARDEN may be able to play a role in triggering conversation topics: "If I visit, there are always dead moments. This will help to get the life back into the conversations."

Creating opportunities through technology. Users with mild cognitive impairment displayed a sense of exploration and adventure when discussing the components of SENSE-GARDEN. The use of virtual environments was perceived as providing ways of visiting new places and experiencing an outdoor environment: "Maybe a place you never went to, but you want to see". One user was particularly impressed with a component of SENSE-GARDEN called "Life Road", which allows the person with dementia to cycle on a stationary bike in front of a film of a familiar place: "We are afraid to ride outside so this option is great. To be safe on a bike." This technology was also considered to provide opportunities for individuals to see places that they can no longer visit: "My mother wants to see her old street again, but we can't do it. With this she can visit again".

These quotes from the users have highlighted the amount of work that goes into creating meaningful experiences for people with dementia. However, with the right kind of facilitation, SENSE-GARDEN may be able to provide these experiences for not only people with dementia, but also for their caregivers.

F. Challenges to Consider

This final theme is perhaps one of the most important in going forward with the SENSE-GARDEN project. The users raised important issues to be aware of when preparing and implementing the SENSE-GARDEN intervention. These concerns are given in the following subthemes: avoiding negative memories, creating personal databases, integrating physical activity, managing symptoms of dementia, and attitudes towards technology.

Avoiding negative memories. Many users emphasised the importance of avoiding stimuli that could evoke negative emotions, such as photographs of relatives who have passed away, for example: "It is necessary to note that there are memories that are not good, and that it is necessary to have very careful prior fieldwork".

Creating personal databases. The main purpose of SENSE-GARDEN is to create experiences that are tailored to the individual with dementia and their past. However, users suggested that there might be challenges in collecting necessary information to achieve this. Issues included the lack of information from family and friends, but potential solutions were also offered by the users: "How you create a database for a lonely person- general triggers in an exploratory approach"..., "The reduced availability of family and friends can be a hindrance. The process of collecting personal data can be eased by using a questionnaire developed for the future SENSE-GARDEN users". A formal caregiver also raised the issue of collecting visual imagery: "We have to see who has videos and films because in this rural area only a few had them".

Integrating physical activity. Whilst many of the respondents emphasised the importance of physical health, issues in implementing physical activity were discussed. One user had a concern regarding the use of a stationary bike for the "Life Road" component of SENSE-GARDEN: "One should reassess the issues of physical activity. For example, bicycles must be those where the person rests and pedals almost lying down".

Managing symptoms of dementia. As in any intervention for people with dementia, it is important to consider how symptoms will be managed and prevented during the sessions: "The person conducting the SENSE-GARDEN session will be essential and must have backup for interventions when disturbed behaviours occur. You don't know how the person will react, even if what you show was a best experience for him". Other issues relating to hallucinations and medication were also raised.

Attitudes towards technology. There was quite a strong sense among the respondents that technology should be hidden during the SENSE-GARDEN sessions: "The experience will be richer when the technology is hidden"..., "The room must be very tempting, persuading- all technology must be hidden". These comments could be interpreted in two ways. Firstly, technology should be hidden to create a more realistic, immersive environment. Contrastingly, it could be that users were referring to the potential reservations that some people have against technology. Some users explicitly expressed negative attitudes towards technology: "Many beware technology". There were also respondents who preferred experiences in natural environments compared to virtual scenarios: "I'd prefer to walk the person with dementia in a real park"…, "SENSE-GARDEN must be just an intermediary step to outdoor and social activities."

This theme has highlighted that whilst the users see many potential benefits for SENSE-GARDEN, they are also aware of the challenges that lay ahead. This affirms the inclusion of not only caregivers, but also people with cognitive impairment in the development of interventions, and the value of adopting a user centred design in interventional research.

V. DISCUSSION

The findings from these user interviews have covered a large variety of ideas regarding the SENSE-GARDEN intervention. Firstly, the respondents were persistent in their beliefs that the environment, the facilitation of the intervention, and the stimuli all need to be tailored to the individual with dementia visiting the SENSE-GARDEN. It should be acknowledged that the task of individualisation is not an easy feat. As human beings, we are all individualistic by nature, with different tastes, preferences, and desires. Adding the constantly fluctuating progression of dementia to this individuality makes it a difficult task in designing technology for these users [21][22]. This is something that the SENSE-GARDEN project will have to tackle through rigorous work and collaboration with users, technology developers, and researchers of various disciplines.

Secondly, the respondents also emphasised the importance of interaction between the SENSE-GARDEN stimuli, the person with dementia, and the caregiver. The respondents` numerous ideas regarding this interaction can be taken forward into a theoretical consideration of technology and its role within SENSE-GARDEN.

A. Technology as the Storyteller: The Potential of Digital Media in Preserving Narrative Identity

Dementia's impact on memory, behaviour and communicative abilities can have detrimental implications for a person's identity. However, there is evidence to suggest that individuals may preserve a sense of self to some extent, even in more severe stages of dementia [19][23]. In this study, there was an overall sense of the immersive environment being able to stimulate autobiographical memory, which was valued as important for preserving a sense of identity. The perspectives of respondents are in agreement with previous research on virtual environments for people with dementia. Siraraya and Ang [24] describe the virtual world as a 'memory sanctuary', in which selfhood and relationships are maintained.

In order to understand how technology and media may be able to preserve identity, we have first to consider what identity means to people with dementia and how it can be shaped by other individuals. The role of others should not be underestimated in maintaining the identity of the person with dementia. In discussing the needs of people with dementia, Kitwood [25] stresses the importance of others in the maintenance of personhood. Westius, Kallenberg, and Norburg [26] present the notion of 'intertwined narrative' in which the life story of the person with dementia is integrated with the narrative of their family carer. Thus, if the person with dementia should become unable to independently recall their story, the intertwined narrative of the caring relationship may provide the opportunity for the maintenance of self.

Earlier literature presents similar ideas. Mills [27] suggests that people with dementia bestow their life stories to another, therefore continuing their sense of identity. Mills states that in this sense, the narrative of the individual never disappears, regardless of the inevitable fading of the person's memory.

One way of preserving this narrative is through the use of digital life books. Digital storytelling, an activity in which technology is used to create innovative forms of narrative, has been shown to educate nursing home staff about the person with dementia [28]. This is especially important for people with dementia living in care homes with little or no family, a challenge mentioned by one of the users included in the present study. Technology and personalised media contents may be the answer to not only preserving, but also sharing that individual's life story with care staff.

SENSE-GARDEN could potentially offer a method for assisting professional caregivers, family and friends in preserving the life story of the person with dementia. Furthermore, the technology of SENSE-GARDEN goes beyond the "life book" concept by offering an entire environment shaped around a person's life. It offers the

opportunity for individuals to become completely immersed in their past. The next section will discuss the technology's place in the overall environment.

B. *Beyond Physical Space: Creating 'Emotional' Environments through a Transactional Relationship*

The respondents emphasised the significance of creating an environment in which the person with dementia and their caregiver could share an emotional experience together. Here, respondents applied meaning beyond the physical space to include emotional and social factors that contribute to the experience of space. In this sense, it is important to have a holistic understanding of what constitutes as an 'environment'.

There is growing acknowledgement of the environment being defined as more than just a physical space. According to literature, an environment is composed of psychosocial elements as well as physical factors [29]. In considering how an environment can shape social interaction, Freund's concept of space is particularly relevant. He writes "space is not merely a place in which social interaction occurs, it structures such interaction" [30].

The way in which an environment simultaneously influences the behaviour of individuals and interpersonal relationships, and yet is shaped by those persons, can be referred to as the transactional relationship. The notion of 'transaction' was firstly used in this context by the philosopher John Dewey, who asserted "Everything that exists in far as it is known and knowable is in interaction with other things. It is associated, as well as solitary, single." [31]. In the context of SENSE-GARDEN, it could be said that a transactional relationship exists between the various technologies (the intervention environment), the person with dementia, and the caregiver. This transactional relationship is conceptualised visually in Figure 2. The figure highlights the numerous interactions that take place between SENSE-GARDEN and its users.

To understand the transactional relationship as a whole, one must consider the individual interactions that take place between each of the three components:

Person with dementia and SENSE-GARDEN stimuli: The SENSE-GARDEN stimuli has a direct effect on the person with dementia, e.g., the system plays a song that evokes a positive reaction in the person with dementia. The SENSE-GARDEN, in turn, will also be influenced by the reactions of the person with dementia. Feedback will enable the system to learn more about the user with each session and therefore, future visits to the SENSE-GARDEN will become increasingly personalised.

Caregiver and SENSE-GARDEN stimuli: The SENSE-GARDEN stimuli may also have an effect on the caregiver. For example, a familiar song might hold significant meaning for an informal caregiver, as well as the person with dementia. In this way, the caregiver may experience their own emotional reaction towards particular stimuli. Alternatively, the caregiver may be indirectly affected by the stimuli through emotional contagion. Emotional contagion refers to the process of an individual's emotional state becoming triggered by emotions displayed in another person

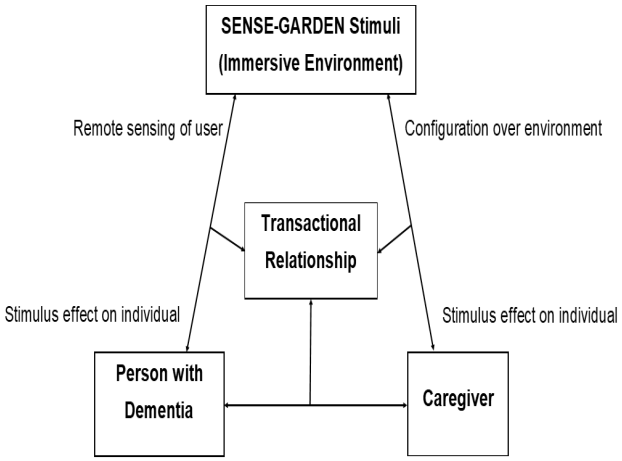


Figure 2. Conceptual model of the transactional relationship that takes place between the person with dementia, caregiver and the SENSE-GARDEN stimuli during the intervention

[32]. In the context of SENSE-GARDEN, the caregiver's emotions may be shaped in response to the reactions of the person with dementia. The caregiver will also be able to configure the SENSE-GARDEN environment based on these reactions, e.g., they can choose to immediately stop a video if it prompts negative behaviour in the person with dementia.

Person with dementia and caregiver: The interpersonal relationship between the person with dementia and the caregiver will shape the entire SENSE-GARDEN experience. For example, if the person with dementia and the caregiver are spouses with a close relationship, they might spend the SENSE-GARDEN session reminiscing on shared moments from their past. However, if the session is taking place between a person with dementia and a new professional caregiver who is not so acquainted with the individual, then their session may involve SENSE-GARDEN providing prompts for the caregiver in order for them to ask questions about the person with dementia's life.

Applying this theoretical frame to the SENSE-GARDEN environment may provide insight into how the intervention works. It will not be possible to gain a full insight into the effects of SENSE-GARDEN without studying the numerous components of the environment. Later literature on emotion echoes Dewey's view, suggesting a need to study the complex relationship between person and environment, for emotions cannot be comprehended by one or the other alone [33].

These ideas can be linked to current thoughts on the nature of technology design, which has been described as "deeply contextual" [22]. Therefore, incorporating the study of context, environment and relationships seems appropriate for both dementia studies and technology development. The interaction between environment and the people within it is vital. How does SENSE-GARDEN, and technology as a whole, fit into this interaction? What role does it play? Going forward, research should adopt a holistic approach to evaluating technology, considering the wider context in which the technology is situated.

VI. Conclusion and Future Work

This paper has demonstrated the value and usefulness of including user groups in the development of not only innovative technologies, but also of interventions for dementia care. Viewing a project through the lens of the user can offer contrasting perspectives with fresh insight into solutions. In the present study, the user interviews yielded valuable insights for the progression of the SENSE-GARDEN project.

The users' value for the relationships within the SENSE-GARDEN suggests that the social and emotional aspects of virtual environments should not be underestimated. This view is supported by previous literature which has called for more research on social interaction in dementia care settings [34]. The results highlight the significance users find in fostering relationships through means of self-identity and emotional relationships. A focus on social and emotional interactions between technology, users, and interpersonal relationships could provide very fruitful results in the context of dementia care. This research provides rationale for the study of emotional engagement and interaction not only in the SENSE-GARDEN project, but also in the wider context of assistive technologies.

The next steps for SENSE-GARDEN include a focus on this emotional aspect. The full trials, planned for summer 2019, will adopt a mixed-methods approach to studying the intervention. Whilst qualitative methods capture rich personal accounts of user experiences, it is important to recognise the value of quantitative measures. Therefore, physiological data will be collected in addition to data from questionnaires, interviews and observation measures. The Empatica E4 wristband [35] will be used to collect information on heart rate and electrodermal activity (EDA). These measures will be assessed during the SENSE-GARDEN visits, as a reaction to different stimuli. Combining this data with qualitative accounts of the SENSE-GARDEN experience will provide a stronger overview of the processes that occur within the intervention.

This research also provides rationale for theoretical work on the role of technology for people with dementia. Firstly, ways of meeting individual needs need to be identified. Respondents stressed the importance of familiarity for the person with dementia, and they raised issues regarding the identification of individual preferences and behaviours.

Secondly, the role of technology as an active contributor to environments, and interactions within those environments, should not go unnoticed. This paper has discussed the SENSE-GARDEN technology in the context of a transactional relationship, but other theories may apply.

Finally, this paper demonstrates the highly interdisciplinary nature of this topic. The users' comments have formed a piece of work that lays at the intersection of human centred design, technology, psychology, sociology, and arguably the creative arts. Future work within dementia care can benefit from incorporating knowledge from these various disciplines.

Acknowledgment

This study was performed in the frame of the EU project SENSE-GARDEN (AAL/Call2016/054-b/2017, with implementation period June 2017 - May 2020) funded by AAL Programme, co-funded by the European Commission and National Funding Agencies of Norway, Belgium, Romania, and Portugal.

We would like to thank all participants in the study.

References

[1] G. Goodall et al., "Using Adaptive Immersive Environments to Stimulate Emotional Expression and Connection in Dementia Care: Insights from User Perspectives towards SENSE-GARDEN", The Fourth International Conference on Human and Social Analytics (HUSO 2018) IARIA, Jun. 2018, pp. 37-42, ISBN: 978-1-61208-648-4

[2] World Health Organization, "Global action plan on the public health response to dementia 2017-2025." Geneva: World Health Organization, 2017.

[3] M. Prince, "World Alzheimer Report 2015: the global impact of dementia: an analysis of prevalence, incidence, costs and trends", Alzheimer's Disease International, 2015.

[4] M. Cadieux, L. Garcia, and J. Patrick, "Needs of People With Dementia in Long-Term Care", American Journal of Alzheimer's Disease & Other Dementias, vol. 28, no. 8, pp. 723-733, Sep. 2013, doi: 10.1177/1533317513500840

[5] R. Milte et al., "Quality in residential care from the perspective of people living with dementia: The importance of personhood", Archives of Gerontology and Geriatrics, vol. 63, pp. 9-17, Mar-Apr, 2016.

[6] K. Lee, D. Algase, and E. McConnell, "Daytime Observed Emotional Expressions of People With Dementia", Nursing Research, vol. 62, no. 4, pp. 218-225, Jul-Aug. 2013, doi: 10.1097/NNR.0b013e31829999d7

[7] E. Finnema, R. Dröes, M. Ribbe, and W. Van Tilburg, "The effects of emotion-oriented approaches in the care for persons suffering from dementia: a review of the literature", International Journal of Geriatric Psychiatry, vol. 15, no. 2, pp. 141-161, 2000.

[8] G. Gowans et al., "Designing a multimedia conversation aid for reminiscence therapy in dementia care environments", in CHI '04 Extended Abstracts on Human Factors in Computing Systems (CHI EA '04), ACM, April. 2004, pp. 825-836, ISBN: 1-58113-703-6.

[9] W. Moyle, C. Jones, T. Dwan, and T. Petrovich, "Effectiveness of a Virtual Reality Forest on People With Dementia: A Mixed Methods Pilot Study", The Gerontologist, vol. 00, no. 00, pp. 1-10, Mar. 2017, doi: 10.1093/geront/gnw270

[10] D. Smit, J. de Lange, B. Willemse, J. Twisk, and A. Pot, "Activity involvement and quality of life of people at different stages of dementia in long term care facilities", Aging & Mental Health, vol. 20, no. 1, pp. 100-109, Jun. 2015, doi: 10.1080/13607863.2015.1049116

[11] Y. Cui, M. Shen, Y. Ma, and S. Wen, "Senses make sense: An individualized multisensory stimulation for dementia", Medical Hypotheses, vol. 98, pp. 11-14, Jan. 2017.

[12] J. Hulsegge and A. Verheul, "Snoezelen: Another World", Chesterfield, UK: Rompa, 2005

[13] R. Orpwood, "Involving People with Dementia in the Design Process- Examples of Iterative Design", in Dementia, Design and Technology, P. Topo and B. Östlund, Ed. IOS Press, 2009, pp. 79-95.

[14] M. Span, M. Hettinga, M. Vernooij-Dassen, J. Eefsting, and C. Smits, "Involving people with dementia in the development of supportive IT applications: A systematic review", Ageing Research Reviews, vol. 12, no. 2, pp. 535-551, Mar. 2013, doi: 10.1016/j.arr.2013.01.002

[15] M. Marshall, "The Needs of People with Dementia and Their Carers and the Potential Role of Design and Technology", in Dementia, Design and Technology, P. Topo and B. Östlund, Ed. IOS Press, 2009, pp. 3-8.

[16] A. Lyon and K. Koerner, "User-Centered Design for Psychosocial Intervention Development and Implementation", Clinical Psychology: Science and Practice, vol. 23, no. 2, pp. 180-200, Jun. 2016, doi: 10.1111/cpsp.12154

[17] V. Braun and V. Clarke, "Using thematic analysis in psychology", Qualitative Research in Psychology, vol. 3, no. 2, pp. 77-101, 2006.

[18] G. Goodall et al., "Dataset for "Using Adaptive Immersive Environments to Stimulate Emotional Expression and Connection in Dementia Care: Insights from User Perspectives towards SENSE-GARDEN", [Data set]. Zenodo. doi:10.5281/zenodo.1232156

[19] C. Surr, "Preservation of self in people with dementia living in residential care: A socio-biographical approach", Social Science & Medicine, vol. 62, no. 7, pp. 1720-1730, 2006

[20] R. Woods, "Discovering the person with Alzheimer's disease: Cognitive, emotional and behavioural aspects", Aging & Mental Health, vol. 5, no. 1, pp. 7-16, 2001.

[21] L. Nygård, "Living with Dementia and the Challenges of Domestic Technology", in Dementia, Design and Technology, P. Topo and B. Östlund, Ed. IOS Press, 2009, pp. 9-25

[22] T. Jiancaro, S. Jaglal, and A. Mihailidis, "Technology, design and dementia: an exploratory survey of developers", Disability and Rehabilitation: Assistive Technology, vol. 12, no. 6, pp. 573-584, Jun. 2016, doi: 10.1080/17483107.2016.1187671

[23] L. Caddell and L. Clare, "The impact of dementia on self and identity: A systematic review", Clinical Psychology Review, vol. 30, no. 1, pp. 113-126, Feb. 2010.

[24] P. Siriaraya and C. S. Ang, "Recreating living experiences from past memories through virtual worlds for people with dementia", SIGCHI Conference on Human Factors in Computing Systems, ACM, Apr. 2014, pp. 3977-3986

[25] T. Kitwood, "The experience of dementia", Aging & Mental Health, vol. 1, no. 1, pp. 13-22, 1997.

[26] A. Westius, K. Kallenberg, and A. Norberg, "Views of life and sense of identity in people with Alzheimer's disease", Ageing and Society, vol. 30, no. 07, pp. 1257-1278, Jul. 2010, doi: 10.1017/S0144686X10000309

[27] M. Mills, "Narrative Identity and Dementia: a Study of Emotion and Narrative in Older People with Dementia", Ageing and Society, vol. 17, no. 6, pp. 673-698, Nov. 1997.

[28] R. Stenhouse, J. Tait, P. Hardy, and T. Sumner, "Dangling conversations: reflections on the process of creating digital stories during a workshop with people with early-stage dementia", Journal of Psychiatric and Mental Health Nursing, vol. 20, no. 2, pp. 134-141, Mar. 2012, doi: 10.1111/j.1365-2850.2012.01900.x

[29] P. Topo and H. Kotilainen, "Designing enabling environments for people with dementia, their family carers and formal carers", in Dementia, design and technology, P. Topo and B. Ostlund, Eds. Amsterdam: IOS Press, pp. 45-60, 2009.

[30] P. Freund, "Bodies, disability and spaces: the social model and disabling spatial organisations," Disability & Society, vol. 16, no.5, pp. 689-706, 2001, doi: 10.1080/09687590120070079

[31] J. Dewey, Experience and Nature. Courier Corporation, 1958.

[32] E. Hatfield, J. T. Cacioppo and R. L. Rapson, "Emotional Contagion", Current Directions in Psychological Science, vol. 2, no. 3, pp. 96-100, Jun. 1993.

[33] R. Lazarus and S. Folkman, "Transactional theory and research on emotions and coping", European Journal of Personality, vol. 1, no. 3, pp. 141-169, Sep. 1987, doi: 10.1002/per.2410010304

[34] J. Mabire, M. Gay, P. Vrignaud, C. Garitte and M. Vernooij-Dassen, "Social interactions between people with dementia: pilot evaluation of an observational instrument in a nursing home," International psychogeriatrics, vol. 28, no.6, pp. 1005-1015, Jun. 2016, doi: 10.1017/S1041610215002483

[35] Empatica. *E4 wristband from empatica.* [Online]. Available from: https://www.empatica.com/en-eu/research/e4/ 2019.05.27

LASSO Regression for Monitoring Patients Progress Following ACL Reconstruction via Motion Sensors: A Case Study

Salvatore Tedesco and Brendan O'Flynn
Tyndall National Institute, University College Cork
Lee Maltings, Prospect Row, Cork, T12R5CP, Ireland
E-mails: salvatore.tedesco@tyndall.ie, brendan.oflynn@tyndall.ie

***Abstract* — Inertial data can represent a rich source of clinically relevant information, which can provide details on motor assessment in subjects undertaking a rehabilitation process. Indeed, in clinical and sport settings, motor assessment is generally conducted through simple subjective measures such as a visual assessment or questionnaire given by caregivers. Thus, inertial sensor technology and associated data sets can help provide an objective and empirical measure of a patient's progress. In this publication, several metrics in different domains have been considered and extrapolated from the three-dimensional accelerometer and angular rate data sets collected on an impaired subject with knee injury, via a wearable sensing system developed at the Tyndall National Institute. These data sets were collected for different activities performed across a number of sessions as the subject progressed through the rehabilitation process. Using these data sets and adopting a combination of techniques (LASSO, elastic net regularization, screening-based approaches, and leave-one-out cross-validation), an automated method has been defined in order to select the most suitable features which could provide accurate quantitative analysis of the improvement of the subject throughout their rehabilitation. The present work confirms that changes in motor ability can be objectively assessed via data-driven methods and that most of the alterations of interest occur on the sagittal plane and may be assessed by an accelerometer worn on the thigh.**

Keywords* — *Regression; Feature Selection; Motor Assessment; Rehabilitation; Wearable.

I. INTRODUCTION

THIS paper is the extended presentation of [1], first published at HealthInfo 2018. While [1] illustrated an effective method for defining a single score indicator which could monitor the rehabilitation progress from features obtained by inertial sensors comparing impaired and unimpaired limb, this work analyses the same features with different data analytics techniques with the goal to investigate which combination of feature, limb, axis and sensor is the most sensitive and helpful to determine changes in motor capacity. Motor assessment is the aspect of biomechanics which studies the process by which the musculoskeletal system can create and control coordinated movements [2]. Voluntary movement requires the transmission of a message from the brain to the appropriate muscle which also controls the smoothness and coordination of the movement. If motor function is intact, muscles can be commanded to move so as to allow symmetrical movements with significant strength levels. However, reduced motor function can occur as a result of injury or trauma to the central nervous system, muscles, ligaments, and so on. Thus, motor impairments can be associated with a number of disorders, such as Parkinson's disease, stroke, cerebral palsy, or orthopeadic injuries, all of them requiring long rehabilitation periods. Therefore, it is essential to track accurately a patient's progress as they proceed through the rehabilitative regimes prescribed to them by their care givers/clinicians, and consequently to tailor patient-specific rehabilitation programs, through the accurate assessment of human motion during the performance of clinically defined tasks, and the development of measured empirical data sets associated with their performance.

With particular reference to the treatment of patients with lower extremity injuries, literature has recently shown a paradigm shift, going from time-dependent concepts to function-based concepts [3], where qualitative and quantitative tests comparing affected and unaffected sides must be met before successfully accessing the following rehabilitation stage.

Qualitative and quantitative motor assessment is typically divided into clinimetrics, balance analysis, and gait analysis.

Indexes, rating scales, questionnaires, and observational forms represent the clinical standard for knee joint assessment, including, for instance, Knee Injury and Osteoarthritis Outcome Score (KOOS), Oxford Knee Score (OKS), Tegner Lysholm Knee Scoring Scale, International Knee Documentation Committee (IKDC), Western Ontario & McMaster Universities Osteoarthritis Index (WOMAC) [4]. However, these tools are subjective and, even when utilised by experienced clinicians, may not be adequate or sensitive enough.

Gold-standard technology adopted in gait analysis for quantitative movement analysis include camera-based motion analysis, instrumented treadmills, force platforms [5], and despite the achieved high performance, their application is constrained by costs, access to specialist motion labs, as well as practicality of application for larger patient/subject groups.

A viable alternative is represented by the adoption of small-size low-cost, wearable sensing units whose consideration for lower-limbs monitoring during rehabilitation, in order to provide objective performance of impaired subjects throughout

the process, has been growing lately. Indeed, inertial sensors, typically including accelerometers, gyroscopes, and magnetometers, have been used to derive gait parameters efficiently both in healthy and symptomatic subjects [6].

As a matter of fact, inertial sensors have been used in a great number of applications, such as navigation systems, activity classification, augmented reality systems, and so on [7][8], and biomechanics, in particular, has achieved significant progress from the adoption of this technology [9].

In the last years, researchers have investigated the possibility to define comprehensive indexes which could quantitatively define gait impairment, thus removing the subjective aspects from the assessment. Some examples are the Gait Deviation Index (GDI) [10], the Gait Profile Score (GPS) [11], and the Classifier Oriented Gait Score (COGS) with related sub-scores [12][13], which provide an indication of the deviation of a subject's conditions in comparison to healthy individuals by taking into account the full-body joint trajectories during walking tasks. However, it may be impractical to measure the full-body joint angles, and similar scores were obtained using limited number of inertial sensors. For instance, Wang et al. [14] defined the Gait Variability Index (GVI) from time-related features obtained by 4 sensors attached on the lower-limbs showing the possibility to monitor gait changes in people with neurodegenerative disorders in a 12 months' period. Likewise, [15] showed that one sensor attached on the lower-back can provide a score (defined as Multifeature Gait Score with related sub-scores considering temporal, symmetry, regularity, complexity, amplitude and distribution aspects) which can assess gait quality by testing the method against healthy adults, sedentary and active older adults. In this case, well-known gait temporal features and time-related acceleration features were adopted for the evaluation.

However, all these examples present some limitation. First of all, these works only consider gait assessments, but do not evaluate additional exercises typically performed during rehabilitation. Moreover, [14][15] which adopted inertial sensors, considered an ageing population of interest and did not test the methods with subjects involved in lower-limbs rehabilitation. Finally, most of these studies considered well-known gait metrics and joint angles for their evaluation. Nevertheless, it has been reported in literature that a data-driven approach may be more beneficial to discriminate impaired from unimpaired subjects. As an example, van den Dikkenberg et al. in [16] observed that, in order to discriminate healthy subjects from Total Knee Replacement (TKR) patients, accelerations (which were significantly different in 213 cases out of 216) were more useful than angles (38 cases out of 52). Furthermore, Patterson et al. [17] studied that gyroscope features were able to discriminate healthy from Anterior Cruciate Ligament (ACL)-reconstructed individuals, which was not possible using spatial or temporal variables.

Most of the studies which apply wearable inertial sensors in lower-limbs rehabilitation generally considered the assessment of impaired subjects against a healthy control during a one-off assessment, or discriminate between correct and incorrect execution in specific rehabilitation exercises [18-21]. However, to date, only a small amount of studies considered the quantitative assessment of patients' performance via inertial sensors during rehabilitation following lower-extremity injuries. This task can be particularly challenging as it consists of isolating the gradual changes in movements due to recovery and improvement despite the presence of a multitude of sources of variability. Indeed, sources of intra- and inter-variability are even more significant in patients following rehabilitation, due to different levels of pain, fatigue, and possible compensations.

Some examples are shown in [22-25]. The main limitations of those studies are related to the short period for data collection which have investigated only the initial part of the rehabilitative process (from 5 days to a maximum of 4 months in [23]). Moreover, the previous works do not evaluate the features extrapolated from the inertial data and their combination so as to show which of them can be the most beneficial and sensitive for clinicians when monitoring patients' movements performed during lower-limb rehabilitation exercises.

The present study analyzes through various data analytic techniques the data collected with the aim of investigating the relationships between inertial-based time-domain features and changes in clinical outcomes and motor performance of adults involved in lower-limb at-home rehabilitation following knee injuries. Besides establishing which of these features are the most sensitive and helpful to determine changes in motor capacity, aspects related to axis, limb, and sensor selection are also investigated, as they could be of relevant importance for reducing the problem dimensionality in the context of wearables where power consumption and computational complexity are of prime concern. The results could help clinicians and sport scientists to gain a comprehensive picture of patients' condition and provide more targeted medical feedback. This investigation is carried out by using a wearable inertial system [26-28] developed at the Tyndall National Institute, consisting of two sensors per limb, able to provide a complete biomechanics assessment for a series of scripted activities.

The present work is organized as follows. Hardware platform description and test protocol are described in Sections II and III, respectively. The features evaluated are illustrated in Section IV. The data analysis is instead performed in Section V. Discussion of the results is illustrated in Section VI. Finally, conclusions are drawn in the final section.

II. HARDWARE PLATFORM

The biomechanical monitoring system consists of two Tyndall Wireless Inertial Measurement Units (WIMUs) per leg [26-28]. The platform measures 44 × 30 × 8 mm and 7.2 g without battery (Figure 1). The WIMU is equipped with a high-performance low-power ARM Cortex-M4 32-bit microprocessor operating at a frequency up to 168 MHz, part of the STM32F0407 family produced by STMicroelectronics. It also features a floating point unit, single precision, high-speed embedded memories (1 Mb of Flash memory, 192 + 4 Kb of SRAM), an extensive range of enhanced I/Os and peripherals, and standard and advanced communication interfaces.

Inertial sensors (three-dimensional accelerometer and gyroscope, MPU-9250 from Invensense) are the main sensing components on the platform and are wired to the microcontroller through the I2C communication. Sensor data can be:

Fig. 1. Tyndall Wireless Inertial Measurement Unit (WIMU).

- transmitted wirelessly via a communication Bluetooth Low-Energy (BLE)-complaint module (Broadcom BCM20737S), representing a single mode low-energy solution with integrated ARM CM3 microcontroller unit, radio frequency and embedded Bluetooth Smart Stack;
- or logged to a removable Micro SD card at 250 Hz.

For measurement of inertial data, the Invensense MPU-9250 was chosen for its low power consumption and the high range (16 g for accelerometer and 2000 deg/s for the gyroscope) with limited noise levels.

The platform also features a USB connector, battery charger, fuel gauge, external I/O connectors, three LEDs, and power switch. All the components were chosen for their specific fitness in mobile applications and, averagely, the overall power consumption in TX/RX mode is 100 mA, dropping to 40 mA (17 mA) for stand-by (sleep mode).

III. Protocol for Data Collection

In conjunction with clinical partners [29], an experimental protocol for data collection was developed to evaluate patient progress. The rehabilitation tasks considered are walking (at defined speeds on a treadmill, e.g., 3, 4, 6 km/h), and exercises such as half-squat, hamstring curl, and flexion-extension, which are defined by physiotherapists as good indicators of rehabilitation progress.

The system has been tested with an impaired subject. The impaired subject is a female athlete, age: 44, height: 161 cm, and weight: 52 kg, with good general health status, with a history of knee injuries and surgery (reconstructed anterior cruciate ligament in the left leg following a sporting injury). The tests were carried out during the course of the rehabilitation program, e.g., starting 1 month before surgery and finishing 7 months after surgery. Overall, the subject has been evaluated in 8 sessions through three periods: once in pre-surgery conditions (e.g., 1 month before surgery), then 6 times in a range of 20 weeks starting one month after surgery (namely short-term post-surgery), and finally once 3 months after the last data capture (e.g., during long-term post-surgery period).

The participant was wearing four devices, two of them were attached to the anterior tibia, 10 cm below the tibial tuberosity, and the remaining two to the lateral thighs, 15 cm above the tibial tuberosity, using surgical adhesive tape.

A number of repetitions have been collected for each exercise, so as to provide an accurate picture of the overall conditions, and each exercise was repeated twice. Most of the exercises were performed during the majority of the data captures. Hamstring curl, as well as walking at 3 and 4 km/h, were performed at every session. Similarly, flexion-extension was always recorded except in the pre-surgery session due to subject's impairment of movement. For the same reason, half-squat and walking at 6 km/h were not recorded in the first 2 sessions after surgery. The order of the exercises within a session was randomized. Prior to participation, the participant received a verbal and written explanation of the study protocol and written consent was obtained. The study received approval by the Clinical Research Ethics Committee at the University College Cork.

IV. Features

The metrics considered for the patient's assessment are well-known statistical features extrapolated from the time-domain. Those variables are applied on every segmented walking stride/exercise repetition for both legs performed during the sessions. More details on the computation of the features are reported in [26]. The selected features are described below:

- Mean, standard deviation, variance, skewness, kurtosis, root mean square (RMS), signal magnitude area, and energy calculated over the acceleration and angular velocity magnitudes,
- Mean, minimum, maximum, median, standard deviation, coefficient of variation (CV), peak-to-peak (p-p) amplitude, and RMS over the x-, y-, and z-axis of the acceleration and angular rate signals.
- Autocorrelation on the x-, y-, and z-axis of the acceleration and angular rate signals measured taking into account all the repetitions/strides in a session as a whole.
- Regularity on the x-, y-, and z-axis of the acceleration and angular rate signals. It is calculated as the ratio between the unbiased autocorrelation coefficient at the first dominant period and the coefficient at the second dominant period, both measured taking into account all the repetitions in a session.

All those features are calculated for both thigh and shank for both legs. Overall, the number of features extracted is $p = 152$, which means that, in this scenario, the number of predictors is much larger than the number of observations n ($p >> n$, $n = 8$).

The data analysis is implemented off-line over the data collected using a commercial software package (MATLAB R2017b, The MathWorks Inc., Natick, MA, 2017).

V. Data Analysis

Preliminary analysis described in [26][27] have highlighted that several parameters are seen to be potentially relevant to provide indications on patient's performance during rehabilitation. However, to support clinicians during their clinical practice, it is essential to understand which of those features are related to currently used clinical indexes.

An accurate assessment of a patient's performance requires the selection of the informative features, excluding those uninformative or redundant. Some features can be informative for some exercises and being redundant for others; thus, it is important to define an automatic method for selecting those

features. A common technique for feature selection is the Least Absolute Shrinkage and Selection Operator (LASSO) [24]. This regression tool requires to define an output in order to adjust the weights of a linear model which defines the features to be selected. However, this method may show some relevant limitations.

In the case of interest, with a number of predictors much larger than the number of observations *n* (also known as, "high dimensional small sample size" - HDSSS), the LASSO can only select *n* variables at most; moreover, if a group of variables is highly correlated, then the LASSO will select only one variable from this group ignoring the others.

These limitations may be overcome with a dual-stage approach. Firstly, screening-based approaches [30][31], which are an effective and computationally efficient method which can reduce the $p >> n$ problem to more acceptable dimensions, could be used to reduce the number of predictors. For example, the Sure Independence Screening (SIS) [32] is a simple method which preserves only those features whose correlation against the responses y_i is above a pre-defined threshold. Secondly, when the number of predictors is strongly reduced, elastic net regularization [33] is adopted to the LASSO, by adding an additional penalty term. The elastic net technique solves the following regularization problem:

$$\min_{\beta_0 \beta} \left(\frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - x_i^T \beta)^2 + \theta P_\alpha(b) \right) \quad (1)$$

Where $P_\alpha(b) = \sum_{j=1}^{p} \left(\frac{1-\alpha}{2} \beta_j^2 + \alpha |\beta_j| \right)$, with y_i being the response at observation *i*, x_i is the *p*-dimensional data, θ is the regularization parameter which controls the strength of the shrinkage of the variables, β a vector of the resulting coefficients of the linear model, and α being the weight of the additional penalty term and included in the range [0-1]. For $\alpha = 0$ the elastic net approaches the ridge regression, while when $\alpha = 1$ it is equivalent to the naïve LASSO technique.

A standard approach to define the regularization parameter when the sample size is small is through leave-one-out cross-validation (LOO CV). A description of this method is shown in [34]. LOO CV is repeated for different values of θ and α calculating, for each pair of coefficients, the related Mean Squared Error (MSE) and R-squared. The performance metrics associated to the minimum-plus-one standard error (1SE) point are then selected. The 1SE point is preferred over the minimum MSE point, as the former usually guarantees to build a model with fewer features. Finally, the optimal regularization parameters related to the models with the minimum MSE among the several models built considering the 1SE distribution are taken into account for feature selection and model generation.

Another aspect to solve in the LASSO approach is provided by the definition of the responses y_i. As shown in [24], this output was defined as linearly increasing from the first to the last test session, with this period ranging from 4 to 12 days. However, even though this assumption can be accepted for the short period of time immediately following surgery, it may be unrealistic when analyzing rehabilitation outcomes for a longer period post-surgery and also pre-surgery.

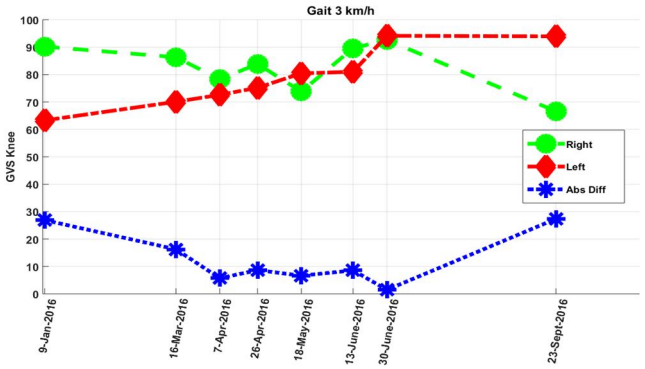


Fig. 2. GVS knee score. Gait 3 km/h.

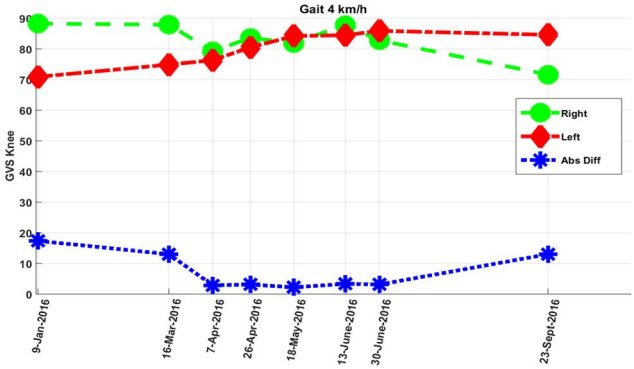


Fig. 3. GVS knee score. Gait 4 km/h.

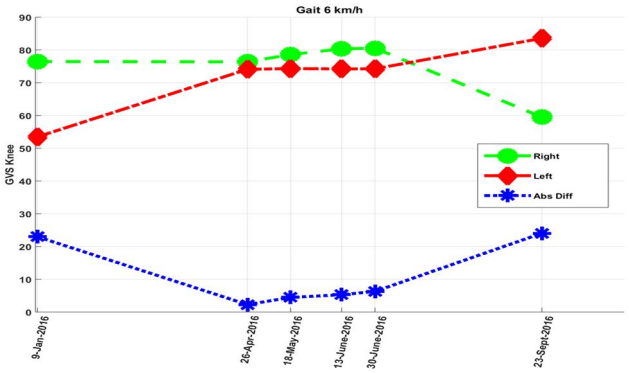


Fig. 4. GVS knee score. Gait 6 km/h.

An alternative may be represented by the adoption of one of the gait indexes discussed in [10-13]; however, those indexes can be based on the joint angles collected from the full-body (as an example, GDI is obtained by taking into account pelvic and hip angles on all three axes, knee flex/extension, ankle dorsi/plantarflexion, and foot progression). Nevertheless, Baker et al. in [11] postulated that those nine variables could be taken individually to calculate a single gait variable, referred to as a Gait Variable Score (GVS), using the same mathematic approach described in [10]. Given that the knee flex/extension angle is obtained from the inertial sensors for all the sessions, the related GVS can be considered as a more accurate option to define the responses y_i. A similar approach, adopted using the Fugl-Meyer and the Wolf Motor Function Test scales in stroke

survivors was studied in [35]. The responses y_i are thus calculated as follows:

1. The knee angle time-normalised curves of all the subjects in a control group are averaged to define a template curve.
2. The natural logarithm of the Euclidean distance between the knee angle time-normalised curve for an individual subject in the control group and the template curve is obtained for all the control group participants, producing a $N_{control}$ *x 1* vector, with $N_{control}$ being the number of subjects in the control group.
3. Mean and standard deviation of the distribution calculated at point 2 are defined.
4. The natural logarithm of the Euclidean distance between the knee angle time-normalised curve for the injured subject in the study and the template curve is obtained.
5. The value calculated at point 4 is standardized using the mean and standard deviation obtained at point 3.
6. The final score (which represents y_i at observation *i*) is calculated by subtracting from 100 the value calculated at point 5 multiplied by 10. As a result, scores of 100 or higher indicate the absence of gait pathologies, while to every 10 points that the score falls below 100 corresponds one standard deviation away from the control group mean.

Normative lower-limb angles data from a control group measured when walking at various speeds were available in [36]. Angles were time-normalised as a percentage of the gait cycle. Given that joint angle in gait is affected by age and gender [37], only data from female subjects with age between 21 and 35 were considered from [36]. As a result, overall 20 subjects were left from the original dataset, with average height equal to 166 cm and average weight of 62 Kg. For the walking speed of 3 km/h, the template curve was obtained from 13 subjects and 75 gait cycles in the speed range of 0.8-1.0 m/s. For the walking speed of 4 km/h, the template curve was obtained from 11 subjects and 61 gait cycles in the speed range of 1.0-1.2 m/s. For the walking speed of 6 km/h, the template curve was obtained from 11 subjects and 53 gait cycles in the speed range of 1.4-1.6 m/s.

It is worth noting that, while the dataset in [36] for the control group has been assembled using the gold-standard VICON as a reference, the knee angles from the injured subject were obtained with the hardware platform described in Section II. Even though this may be problematic, it has been demonstrated in [28] that this platform guarantees an average error in the estimation of the knee angle equal to -0.29, 1.54 and 1.58 deg at 3, 4, and 6 km/h, respectively (or 5.2, 7.4, and 11.3 deg if considering the RMS error), and thus it was deemed comparable with the gold-standard technology. The gait cycles were automatically segmented from the motion data using the procedure described in [38], while the joint angles were estimated via the algorithm in [39].

Examples of the knee angle-based GVS for left and right leg over the eight sessions for the various speeds are shown in Figures 2-4, together with the related absolute differences.

From the figures, it is evident how the left leg shows an almost linear improvement in the GVS score over the different sessions. While the score increment stopped at the 7th session for walking at 3 and 4 km/h, it continued up to the 8th session for 6 km/h speed. On the other hand, for the first 7 sessions, the right leg tends to have a constant score, with limited variability at the increase of the speed. However, the last session shows a substantial drop in performance in the right leg, and this is evident on all the walking speeds. As a result, the absolute difference between the legs shows a linear almost monotonic trend tending towards zero, which confirms similar results in literature [40-42], with however a large value in the last session. Unfortunately, authors are not aware of a reasonable explanation for the shown behavior, which could be due to excessive training loads, fatigue, movement compensation dysfunction, etc. and since it occurs in the last session, it is not feasible to indicate it as an outlier or not. As a consequence, the GVS score reported is used as the responses y_i in the LASSO in two cases, with and without considering the last session.

In summary, the described data analysis includes the following steps:

1. From the collected raw motion data, each repetition/cycle is segmented and the related joint angle is estimated;
2. Using datasets available online with normative data, the GVS knee score for the walking tasks is defined for each session which corresponds to the responses y_i in the LASSO technique;
3. From the collected raw data, the features described in Section II are extrapolated for both legs, and the related absolute difference is obtained, which is afterwards standardized considering all the observations;
4. Relying on the SIS method, the features with a correlation against y_i lower than 0.7 are discarded;
5. Using the LOO CV approach, the LASSO problem is solved defining the best θ and α coefficients, and thus the related selected features and β vectors.

VI. RESULTS

In each session, each exercise was divided in two separate tests (both logged for 60 sec), and in each of the two tests a series of repetitions have been carried out by the subject. The overall number of repetitions recorded for all the sessions was: 184 hamstring curls (92 left / 92 right), 134 flexion/extensions (67 left / 67 right), 66 half squats, 478 strides for both legs when walking at 3 km/h, and similarly 544 strides when walking at 4 km/h, and 512 strides when walking at 6 km/h.

For each test, the features described in Section IV, were extrapolated and compared among the different sessions after applying the data analysis defined in Section V.

Owing to technical issues with system hardware during data recording, data from the right leg in the hamstring curl exercise on the first session are not available.

WIMUs have been attached to the anterior tibia, 10 cm

below the tibial tuberosity, and to the lateral thigh, 15 cm above the tibial tuberosity using surgical adhesive tape.

Finally, in order to have the same reference system for both WIMUs worn on the same leg, the method proposed by Seel et al. [43] has been adopted to virtually rotate around an axis the raw inertial data recorded on the shank. As a result, for all the WIMUs involved, the x-axis represents the mediolateral axis, the y-axis is the anteroposterior one, while the z-axis is the vertical axis. Thus, the plane y-z represents the sagittal plane. Results for all the exercises are described below.

A. Gait 3 km/h

Considering the gait task at 3 km/h, the resulting selected features are summarized in Table I.

As expected, when considering the responses y_i without the last session, better results in terms of MSE and R squared are achieved (MSE = 12.62 and R^2 = 0.98 vs. MSE = 16.82 and R^2 = 0.92 if the last session is kept). As a consequence, the model built without removing the 8th observation is simpler as fewer features show this particular behaviour. This model consists of 3 features versus 9 features for the model without the possible outlier. However, 2 out of the 3 features are in common between the two models. The number of features kept after applying SIS was 9 and 15, respectively.

B. Gait 4 km/h

Considering the gait task at 4 km/h, the resulting selected features are summarized in Table II.

Unlike the previous case, considering the last session achieves a model with better performance metrics in terms of MSE and R-squared despite having twice the number of metrics. Interestingly, all the features in the model built without the last session are included in the model built considering all the sessions. The number of features kept after applying SIS was 18 and 25, respectively.

C. Gait 6 km/h

Considering the gait task at 6 km/h, the resulting selected features are summarized in Table III.

Again, considering all the sessions provides the best performance metrics in terms of MSE (6.18) and R-squared (0.96) selecting 29 features instead of the 44 chosen by the model that does not consider the last session. 21 out of the 29 features are included in both models. The number of features kept after applying SIS was 29 and 44, respectively.

D. Other Tasks

To the best of authors' knowledge, clinical indexes and ratings are available in literature only for gait tasks. Addressing these aspects also for general exercises in a rehabilitation process would be an important aspect in order to empower clinicians in their clinical practice. Given that the GVS score for the knee joint was similar at every speed, an average of those scores has been considered as an indication of the responses y_i when analyzing other non-walking tasks/exercises, as a better alternative than simply using a linear model.

For the flexion/extension exercise, the resulting selected features are summarized in Table IV.

Interestingly, in the model built without the last session, no features were selected with the best performance provided by a simple constant linear model; however, considering all the sessions suggested a more interesting model with good MSE/R-squared performance consisting of a limited number of features due to the chosen naïve LASSO approach. The number of features kept after applying SIS was 20 and 38, respectively.

For the hamstring curl exercise, the resulting selected features are summarized in Table V.

Again, the model built without the last session selected no features while the model considering all the sessions suggested to select 14 features. The number of features kept after applying SIS was 14 and 20, respectively.

For the half-squat exercise, the resulting selected features are summarized in Table VI.

The model built considering all the sessions provide the best results in terms of MSE and R-squared (7.51 and 0.98, respectively), together with a reduced number of features selected compared to the other model. Interestingly, 23 out of the 28 overlap between the two models. The number of features kept after applying SIS was 28 and 44, respectively.

E. Discussion

ACL injuries are common and functionally disabling. The biomechanical effects of ACL injuries are well-known in literature; however, few studies have followed individuals throughout the whole rehabilitation and treatment period. As shown in [44], alterations in frontal- and sagittal-plane walking kinematics and kinetics observed early (< 12 months) after surgery persisted in the following period (12-36 months). Despite clearance to return to physical activity, these gait patterns do not appear to normalize over time, which may indicate that the current approach to rehabilitation and assessment before return to activity is not adequately identifying individuals with dysfunctional movement patterns. This was also confirmed in [45], where biomechanical differences between limbs were observed 9 months after reconstruction across jump/landing tasks, and in [42] where joint kinematics differences were observed up to 6 years following reconstruction.

Therefore, a data-driven approach may be more suitable in order to identify those dysfunctional movement patterns during the rehabilitation process. Inertial sensors can then bring a huge impact on clinical practice.

This work analyzed the body-worn inertial data collected from a patient over the course of rehabilitation adopting a combination of techniques (LASSO, elastic net regularization, SIS, LOO CV, and quantitative clinical indexes) for the definition of an automated method which could select a number of features for better understanding and monitoring patient's progress in several test and, thus, predict the clinical outcome.

The resulting performance, for models considering all the therapeutic sessions, shows good values in terms of both MSE (average result 10.8) and R-squared (going from 0.9025 to 0.98). A summary of the features selected is shown in Table VII.

From the table, it is evident that accelerometer and gyroscope-derived features have the same importance in a

TABLE I. GAIT 3 KM/H

	Features selected	*MSE*	R^2	α
With all sessions	Accelerometer thigh (1): Mean Y-axis, Gyroscope thigh (1): Mean Z-axis, Gyroscope shank (1): Maximum Y-axis	16.82	0.928	1
Without the last session	Accelerometer thigh (3): Mean Y-axis, Median Y-axis, CV X-axis Gyroscope thigh (1): Mean Z-axis Gyroscope shank (5): RMS magnitude, Mean X-axis, Minimum Y-axis, CV Y-axis, p-p amplitude Y-axis	12.62	0.984	0.8

TABLE II. GAIT 4 KM/H

	Features selected	*MSE*	R^2	α
With all sessions	Accelerometer thigh (2): Mean Y-axis, Median Y-axis Accelerometer shank (3): RMS magnitude, RMS X-axis, Regularity Z-axis Gyroscope thigh (3): Maximum X-axis, st_dev X-axis, p-p amplitude X-axis Gyroscope shank (2): Maximum X-axis, RMS X-axis	11.26	0.9025	0.7
Without the last session	Accelerometer thigh (2): Mean Y-axis, Median Y-axis Gyroscope thigh (2): Maximum X-axis, St_dev X-axis Gyroscope shank (1): Maximum X-axis	15.78	0.717	0.9

TABLE III. GAIT 6 KM/H

	Features selected	*MSE*	R^2	α
With all sessions	Accelerometer thigh (14): St_dev/Variance/Skewness magnitude, Mean X-Y-Z-axis, Minimum Y-Z-axis, Maximum Y-axis, Median X-Y-Z axis, autocorrelation Y-axis, Regularity Y-axis Accelerometer shank (2): CV X-axis, p-p amplitude Y-axis Gyroscope thigh (10): Mean/Skewness/RMS/Area/Energy magnitude, Maximum Z-axis, Median Z-axis, St_dev X-axis, p-p amplitude Y-Z-axis, Gyroscope shank (3): Median Y-axis, St_dev Z-axis, p-p amplitude Z-axis	6.19	0.96	0.3
Without the last session	Accelerometer thigh (17): St_dev/Variance magnitude, Mean X-Y-axis, Minimum X-Y-axis, Maximum Y-axis, Median X-Y-Z-axis, St_dev Z-axis, CV X-Y-Z-axis, RMS X-axis, autocorrelation Y-axis, Regularity Y-axis Accelerometer shank (4): Maximum X-axis, CV X-axis, p-p amplitude X-Y-axis Gyroscope thigh (14): Mean/Skewness/Area/Energy magnitude, Mean Z-axis, Maximum X-Z-axis, Median Z-axis, St_dev X-axis, CV X-Y-axis, p-p amplitude X-axis, RMS X-axis, autocorrelation Z-axis	63.7	0.53	0.2
	Gyroscope shank (9): Mean X-Z-axis, Minimum X-axis, Maximum X-axis, Median Y-axis, St_dev X-axis, p-p amplitude X-axis, autocorrelation Z-axis, Regularity Z-axis			

TABLE IV. FLEXION/EXTENSION TASK

	Features selected	*MSE*	R^2	α
With all sessions	Accelerometer thigh (1): Mean Y-axis Gyroscope shank (1): Mean Z-axis	6.04	0.95	1
Without the last session	N/A	21.35	0	0.9

TABLE V. HAMSTRING CURL TASK

	Features selected	*MSE*	R^2	α
With all sessions	Accelerometer thigh (7): Mean Y-axis, Minimum Y-axis, Maximum Y-axis, Median Y-Z-axis, RMS Y-axis, Regularity Y-axis Accelerometer shank (5): Skewness/Area/Energy magnitude, autocorrelation Z-axis, Regularity X-axis Gyroscope thigh (1): RMS Y-axis Gyroscope shank (1): Mean Y-axis	16.98	0.93	0.1
Without the last session	N/A	21.28	0	0.9

TABLE VI. HALF-SQUAT TASK

	Features selected	*MSE*	R^2	α
With all sessions	Accelerometer thigh (18): Variance magnitude, Mean X-Y-Z-axis, Minimum X-Y-Z-axis, Maximum Z-axis, Median X-Y-Z-axis, St_dev Z-axis, CV Y-Z-axis, p-p amplitude Z-axis, RMS Y-Z-axis, Regularity Z-axis Accelerometer shank (4): Maximum X-axis, p-p amplitude X-Y-axis, autocorrelation Z-axis Gyroscope thigh (4): Median X-axis, RMS Y-Z-axis, autocorrelation X-axis Gyroscope shank (2): RMS magnitude, CV Z-axis	7.51	0.98	0.1
Without the last session	Accelerometer thigh (20): Mean X-Y-Z-axis, Minimum X-Y-Z-axis, Maximum X-Y-Z-axis, Median X-Y-Z-axis, St_dev Z-axis, CV Y-Z-axis, p-p amplitude Z-axis, RMS Y-Z-axis, Regularity Y-Z-axis Accelerometer shank (6): Mean/Skewness/Area magnitude, Maximum X-axis, p-p amplitude Y-axis, Regularity Z-axis Gyroscope thigh (5): Maximum X-Y-axis, RMS Y-Z-axis, autocorrelation Z-axis Gyroscope shank (8): Mean/Skewness/RMS/Area/Energy magnitude, Maximum X-axis, CV X-Z-axis	77.69	0.65	0.3

TABLE VII. FEATURES SELECTED - SUMMARY

		Gait 3 km/h	*Gait 4 km/h*	*Gait 6 km/h*	*Flex/Extension*	*Hamstring Curl*	*Half-Squat*	*Total*
Thigh	Accelerometer magnitude			3			1	4
	Accelerometer X-axis			2			3	5
	Accelerometer Y-axis	1	2	6	1	6	5	21
	Accelerometer Z-axis			3		1	9	13
	Gyroscope magnitude			5				5
	Gyroscope X-axis		3	1			2	6
	Gyroscope Y-axis			1		1	1	3
	Gyroscope Z-axis	1		3			1	5
Shank	Accelerometer magnitude		1			3		4
	Accelerometer X-axis		1	1		1	2	5
	Accelerometer Y-axis			1			1	2
	Accelerometer Z-axis		1			1	1	3
	Gyroscope magnitude						1	1
	Gyroscope X-axis		2					2
	Gyroscope Y-axis	1		1		1		3
	Gyroscope Z-axis			2	1		1	4
	Total	3	10	29	2	14	28	86
Sensors	Accelerometer	1	5	16	1	12	22	57
	Gyroscope	2	5	13	1	2	6	29
Limbs	Thigh	2	5	24	1	8	22	62
	Shank	1	5	5	1	6	6	24

number of tasks, such as walking at different speeds and flexion/extension. However, this is not shown for other exercises, e.g., hamstring curl and half-squat, where features selected from the accelerometer are present in a larger number. This may indicate that accelerometry may be sufficient to detect incorrect movement patterns, and this can be even more important in battery-powered devices, considering the gyroscope power consumption is typically larger than the accelerometer's (almost 7 times larger in the platform built in Section II). Secondly, features obtained from thigh and shank limbs have similar distributions in a number of tasks (walking at 3-4 km/h, flexion/extension, and hamstring curl), except for gait at 6 km/h and half-squat, where thigh-derived features are more prominent. This can be explained by the fact that those two exercises are more physically demanding for the subjects, and ACL tears causes a decrement in the quadriceps and hamstring muscles, with the decrease in quadriceps strength being 3-fold greater [46]. Thus, a limited strength in the thigh muscles can limit the control of the knee and lower limbs during complex and demanding activities in the rehabilitation phase. This aspect may be further investigated by adding electromyography (EMG) sensors in future analysis. Finally, when taking into account the individual features, it is evident how features extrapolated from the accelerometer over the anteroposterior axis on the thigh are the only features present in every task. Features from the other axis, sensors, and limbs are uniformly distributed across the different exercises, with no preference of one over the other, except for the accelerometer-based features extrapolated from the vertical axis of the thigh, which are found in walking at 6 km/h and half-squat tasks. Over 24% of the overall features selected in all the tasks is obtained from the anteroposterior axis of the accelerometer located on the thigh, and 15% of the overall features are from the vertical axis. These findings confirm the results discussed in [44], with most of the alterations of interest taking place in the sagittal plane.

Regarding the application of the SIS method, the number of features kept after using this method was included between 9 and 29 (when considering all the therapeutic sessions), and between 15 and 44 when not considering the last session. In percentage, the kept features represent a fraction of the originally considered features in the range of 5.9-19% (with all sessions), and 9.8-28.9% (without the last session). This confirms the effectiveness of the SIS method in reducing $p >> n$ problems to more acceptable dimensions. Another interesting aspect is that the number of features kept with SIS in the model built with all the sessions is lower than the number of features kept without considering the last session for all the exercises. This was expected since the last session introduces an unexpected behavior in the responses y_i which is more unlikely to be reproduced in the analyzed features.

Finally, it is also worth investigating that not always all the features kept using SIS are then adopted to build a model with the following LASSO/elastic net regularization. In fact,

exercises such as walking at 6 km/h, hamstring curl, and squat, show a model defined with the exact number of features filtered with SIS; this was expected as in those models the parameter α in the elastic net regularization approach is between 0.1 and 0.3, thus approaching a ridge regression which does not have the ability to further reduce the number of features. On the other side, exercises such as walking at 3 and 4 km/h and flexion/extension, show a model defined using between 10 and 55% of the features filtered with SIS; this occurs as in those models the parameter *a* is included between 0.7 and 1, thus approaching the naïve LASSO technique which has the ability to further reduce the number of features.

This study only considered well-known time-domain statistical features for this analysis, extrapolated from acceleration and angular rate signals of the shank and thigh, proving their sensitivity for a number of exercises. However, as only a single subject was available for the present study, an enhanced number of athletes, with homogeneous characteristics, will also be tested to have a more robust base and further validate the drawn conclusions.

VII. Conclusion

This work presented a combination of wearable inertial-based system and data analytics techniques for an objective assessment of lower-limbs in patients over the course of rehabilitation. The hardware platform adopted for the system realization and the data analytics involving inertial data collected from thighs and shanks have been described.

The studied techniques are able to indicate which features are more informative regarding patients' performance and could be easily taken into account by clinicians during their analysis. Results analysis confirmed that changes in motor ability can be objectively assessed via data-driven methods and that most of the alterations of interest occur on the sagittal plane and may be assessed by an accelerometer worn on the thigh. Future work should further assess the system capability to differentiate injured and non-injured subjects collecting larger datasets by recruiting a greater amount of participants and involving more exercise types. Moreover, datasets could be enriched by including additional sensing technologies, e.g., EMG, galvanic skin response, heart rate. Despite the availability of a number of public datasets, a rehabilitation dataset including the described characteristics is not yet available in literature and would further develop this area. Therefore, additional clinical trials are currently being planned in order to further validate the developed model in statistical terms. Moreover, the development of personalized models could be further investigated also adopting different data analytics methods, such as deep learning techniques.

Acknowledgments

This publication has emanated from research supported by a research grant from Enterprise Ireland (EI) under grant number CF-2017-0685-P. Aspects of this work have emanated in part from research supported in part by Science Foundation Ireland (SFI) and is co-funded under the European Regional Development Fund, Grant Number 13/RC/2077-CONNECT.

The authors would like to thank Sebastian Scheurer for the insightful suggestions and comments.

References

[1] B. O'Flynn, N. Dagini, S. Bottone, A. Urru, and S. Tedesco, "Metrics for monitoring patients progress in a rehabilitation context," *Int Conf Healthinfo* 2018, pp. 42-47.

[2] P. D. Hansen and L. S. Chandler, "Assessment of sensory and motor function," in *Rehabilitation*, Human Brain Function (Assessment and Rehabilitation), G. Goldstein, S. R. Beers, Springer, Boston, MA, USA, 1998, pp. 131-153.

[3] M. Keller, E. Kurz, O. Schmidtlein, G. Welsch, and C. Anders, "Interdisciplinary assessment criteria for rehabilitation after injuries of the lower extremity: A function-based return to activity program," *Sportverletz Sportshaden.*, vol. 30, no. 1, pp. 38-49, 2016.

[4] E. C. Rodriguez-Merchan, "Knee instruments and rating scales designed to measure outcomes," *J Orthop Traumatol*, vol. 13, no. 1, pp. 1-6, 2012.

[5] G. Robertson, G. Caldwell, J. Hamill, G. Kamen, and S. Whittlesey, *Research Methods in Biomechanics, Second Edition*, Human Kinetics, Champaign, IL, USA: 2014.

[6] C. M. Kanzler, J. Barth, J. Klucken, and B. M. Eskofier, "Inertial sensor based gait analysis discriminates subjects with and without visual impairment caused by simulated macular degeneration," *Proc IEEE Eng Med Biol Soc*, Orlando, FL, USA, Aug 16-20, 2016, pp. 4979-4982.

[7] M. Walsh, S. Tedesco, T.C. Ye, and B. O'Flynn, "A wearable hybrid IEEE 802.15.4a ultra-wideband/inertial sensor platform for ambulatory tracking," *Proc IEEE Int Conf Body Area Networks (BodyNets)*, London, UK, Sept 29-Oct 1, 2014, pp. 352-357.

[8] S. Scheurer, S. Tedesco, K.N. Brown, and B. O'Flynn, "Human activity recognition for emergency first responders via body-worn inertial sensors," *Proc IEEE Int Conf Wearable and Implantable Body Sensor Networks (BSN)*, Eindhoven, NL, May 9-12, 2017, pp. 5-8.

[9] B. O'Flynn, J. Torres Sanchez, S. Tedesco, B. Downes, J. Connolly, J. Condell, and K. Curran, "Novel smart glove technology as a biomechanical monitoring tool," *Sensors & Transducers Journal*, Vol. 193, No. 10, pp. 23-32, 2015.

[10] M. H. Schwartz and A. Rozumalski, "The gait deviation index: A new comprehensive index of gait pathology," *Gait & Posture*, vol. 28, pp. 351-357, 2008.

[11] R. Baker et al., "The gait profile score and movement analysis profile," *Gait & Posture*, vol. 30, pp. 265-269, 2009.

[12] J. Christian, J. Kroll, G. Strutzenberger, N. Alexander, M. Ofner, and H. Schwameder, "Computer aided analysis of gait patterns in patients with acute anterior cruciate ligament injury," *Clinical Biomechanics*, vol. 33, pp. 55-60, 2016.

[13] J. Christian, J. Kroll, and H. Schwameder, "Comparison of the classifier oriented gait score and the gait profile score based on imitated gait impairments," *Gait & Posture*, vol. 55, pp. 49-54, 2017.

[14] X. Wang, D. Ristic-Durrant, M. Spranger, and A. Graser, "Gait assessment system based on novel gait variability measures," *Proc Int Conf Rehab Robotics (ICORR)*, London, UK, July 17-20, 2017, pp. 467-472.

[15] K. Ben Mansour, P. Gorce, and N. Rezzoug, "The multifeature gait score: An accurate way to assess gait quality," *PLoS ONE*, vol. 12, no. 10, e0185741, 2017.

[16] N. van den Dikkenberg et al., "Measuring functional abilities of patients with knee problems: rational and construction of the DynaPort knee test," *Knee Surg, Sports Traumatol, Arthrosc*, vol. 10, pp. 204-212, 2002.

[17] M. R. Patterson, E. Delahunt, K. T. Sweeney, and B. Caulfield, "An ambulatory method of identifying anterior cruciate ligament reconstructed gait patterns," *Sensors*, vol. 14, 99. 887-899, 2014.

[18] O. M. Giggins, K. T. Sweeney, and B. Caulfield, "Rehabilitation exercise assessment using inertial sensors: a cross-sectional analytical study," *J NeuroEng Rehab*, vol. 11, pp. 158, 2014.

[19] E. Papi, D. Osei-Kuffour, Y. M. A. Chen, and A. H. McGregor, "Use of wearable technology for performance assessment: A validation study," *Med Eng Phys*, vol. 37, no. 7, pp. 698-704, 2015.

[20] K. H. Chen, P. C. Chen, K. C. Liu, and C. T. Chan, "Wearable sensor-based rehabilitation exercise assessment for knee osteoarthritis," *Sensors*, vol. 15, no. 2, pp. 4193-4211, 2015.

[21] V. Bonnet, V. Joukov, D. Kulic, P. Fraisse, N. Ramdani, and G. Venture, "Monitoring of hip and knee joint angles using a single inertial measurement unit during lower limb rehabilitation," *IEEE Sensors Journal*, vol. 16, no. 6, pp.1557-1564, 2016.

[22] J. F. S. Lin and D. Kulic, "Human pose recovery for rehabilitation using ambulatory sensors," *Proc IEEE Eng Med Biol*, Osaka, Japan, July 3-7, 2013, pp. 4799-4802.

[23] M. Field, D. Stirling, M. Ros, Z. Pan, and F. Naghdy, "Inertial sensing for human motor control symmetry in injury rehabilitation," *Proc IEEE/ASME Advanced Intelligent Mechatronics*, Wollongong, NSW, Australia, July 9-12, 2013, pp. 1470-1475.

[24] R. Houmanfar, M. Karg, and D. Kulic, "Movement analysis of rehabilitation exercises: distance metrics for measuring patient progress," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1014-1025, 2016.

[25] C. Y. Chiang, K. H. Chen, K. C. Liu, S. J. P. Hsu, and C. T. Chan, "Data collection and analysis using wearable sensors for monitoring knee range of motion after total knee arthroplasty," *Sensors*, vol. 17, no. 2, 418, 2017.

[26] S. Tedesco, A. Urru, J. Peckitt, and B. O'Flynn, "Inertial sensors-based lower-limb rehabilitation assessment: A comprehensive evaluation of gait, kinematic and statistical metrics," *Int J Advances Life Sciences*, vol. 9, no. 1-2, pp. 33-49, 2017.

[27] S. Tedesco, A. Urru, and B. O'Flynn, "Spectral and time-frequency domains features for quantitative lower-limb rehabilitation monitoring via wearable inertial sensors," *Proc IEEE Bio Circuits Systems*, Turin, Italy, Oct 19-21, 2017.

[28] S. Tedesco, A. Urru, A. Clifford, and B. O'Flynn, "Experimental validation of the Tyndall portable lower-limb analysis system with wearable inertial sensors," *Procedia Engineering: 11th Conf Int Sport Eng Assoc, ISEA, The Engineering of Sport 11*, vol. 147, pp. 208-213, 2016.

[29] P. Ford, E. Coughlan, and M. Williams, "The expert-performance approach as a framework for understanding and enhancing coaching performance, expertise and learning," *International Journal of Sports Science & Coaching*, vol. 4, no. 3, pp. 451-463, 2009.

[30] J. Fan and J. Lv, "A selective overview of variable selection in high dimensional feature space," *Stat Sin*, vol. 20, no. 1, pp. 101-148, 2010.

[31] L.D.D. Desboulets, "A review on variable selection in regression analysis," *Econometrics*, vol. 6, no. 45, 2018.

[32] J. Fan and J. Lv, "Sure independence screening for ultra-high dimensional feature space," *Journal of the Royal Statistical Society, Series B*, vol. 70, no. 5, pp. 849-911, 2008.

[33] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society, Series B*, vol. 67, no. 2, pp. 301–320, 2005.

[34] W.H. Finch, and M.E.H. Finch, "Regularization methods for fitting linear models with small sample sizes: Fitting the lasso estimator using R," *Practical Assessment, Research & Evaluation*, vol. 21, no. 7, 2016.

[35] Y.A. Majeed, S.S. Awadalla, and J.L Patton, "Regression techniques employing feature selection to predict clinical outcomes in stroke," *PLoS ONE*, vol. 13, no. 10, e0205639, 2018.

[36] B.F. Mentiplay, M. Banky, R.A. Clark, M.B. Kahn, and G. Williams, "Lower limb angular velocity during walking at various speeds," *Gait & Posture*, vol. 65, pp. 190-196, 2018.

[37] T. Oberg, A. Karsznia, and K. Oberg, "Joint angle parameters in gait: Reference data for normal subjects, 10-79 years of age," *Journal of Rehabilitation Research and Development*, vol. 31, no. 3, pp. 199-213, 1994.

[38] D. McGrath, B.R. Greene, K.J. O'Donovan, and B. Caulfield, "Gyroscope-based assessment of temporal gait parameters during treadmill walking and running," *Sports Eng*, vol. 15, pp. 207-213, 2012.

[39] S.O.H. Madgwick, A.J.L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," *Proc IEEE Int Conf Rehabilitation Robotics*, Zurich, 2011.

[40] S.D. Tsivgoulis et al., "Pre-operative versus post-operative gait variability in patients with acute anterior cruciate ligament deficiency," *The Journal of International Medical Research*, vol. 39, pp. 580-593, 2011.

[41] A. Alghadir, Z.A. Iqbal, and S. Anwer, "Comparison of the effect of pre- and post-operative physical therapy versus post-operative physical therapy alone on pain and recovery of function after total knee arthroplasty," *The Journal of Physical Therapy Science*, vol. 28, pp. 2754-2758, 2016.

[42] M. Kaur, D.C. Ribeiro, J.C. Theis, K.E. Webster, and G. Sole, "Movement patterns of the knee during gait following ACL reconstruction: A systematic review and meta-analysis," *Sports Med*, vol. 46, no. 12, pp. 1869-1895, 2016.

[43] T. Seel, J. Raisch, and T. Schauer, "IMU-based joint angle measurement for gait analysis," *Sensors*, vol. 14, no. 4, pp. 6891-6909, 2016.

[44] L.V. Slater, J.M. Hart, A.R. Kelly, and C.M. Kuenze, "Progressive changes in walking kinematics and kinetics after anterior cruciate ligament injury and reconstruction: A review and meta-analysis," *Journal of Athletic Training*, vol. 52, no. 9, pp. 847-860, 2017.

[45] E. King et al., "Whole body biomechanical differences between limbs exist 9 months after ACL reconstruction across jump/landing tasks," *Scandinavian Journal of Medicine and Science in Sports*, vol. 28, no. 12, pp. 2567-2578, 2018.

[46] H.J. Kim, J.H. Lee, S.E. Ahn, M.J. Park, and D.H. Lee, "Influence of anterior cruciate ligament tear on thigh muscle strength and hamstring-to-quadriceps ratio: A meta-analysis," *PLoS One*, vol. 11, no. 1, e0146234, 2016.

Knowledge Distillation from Machine Learning Models for Prediction of Hemodialysis Outcomes

Harry Freitas da Cruz, Siegfried Horschig and Matthieu-P. Schapranow

Digital Health Center
Hasso Plattner Institute
Rudolf-Breitscheid-Str. 187, 14482 Potsdam, Germany
Email: {harry.freitasdacruz|schapranow}@hpi.de
siegfried.horschig@student.hpi.de

Christian Nusshag

Department of Nephrology
Heidelberg University Hospital
Im Neuenheimer Feld 162, 69120 Heidelberg, Germany
Email: christian.nusshag@med.uni-heidelberg.de

***Abstract*—In order to compensate severe impairments of renal function, artificial, extracorporeal devices, so called dialyzers, have been developed to enable renal replacement therapy. The parameters utilized in this form of therapy and the specific patient characteristics substantially affect individual patient outcomes and overall disease progression. In this paper, we present a clinical prediction model for outcomes of critically ill patients that underwent a specific form of renal replacement, hemodialysis. For this purpose, we employed two categories of machine learning models: interpretable (Bayesian rule lists and logistic regression) and non-interpretable (multilayer perceptron and random forest). To provide more transparency to the latter category, we applied mimic learning and feature importance metrics. Results show that non-interpretable models outperform the rule-based classifier (c-statistic $\geq$ 0.9). Despite this result, the use of interpretability methods enables more thorough model scrutiny by a medical experts, revealing possible model biases, which might have been otherwise disregarded.**

Keywords–clinical prediction model; renal replacement therapy; machine learning; supervised learning; knowledge distillation.

I. INTRODUCTION

Previously, we developed a prediction model for patient outcomes following Renal Replacement Therapy (RRT) [1]. In this paper, we expand our previous work, including different algorithms, metrics, and more in-depth discussion so as to provide a more comprehensive picture of the contributions, challenges and limitations faced.

The renal system in the human body has the purpose to excrete predominantly water-soluble metabolites and toxins in order to maintain a sufficient blood homeostasis [2]. If this system is impaired severely, e.g., in the context of an Acute Kidney Injury (AKI), artificial, extracorporeal organ replacement therapy becomes necessary [3]. Therefore, different RRT modalities are available. One example is the hemodialysis, where the solute exchange takes place via diffusion across a semipermeable membrane between the blood and the dialysate or dialysis fluid [4].

Hemodialysis outcomes are highly dependent on the patient's clinical characteristics as well as on the type of the RRT procedure applied [5]. Furthermore, RRT modalities based on a filtration circuit, such as hemofiltration or hemodiafiltration are particularly costly, requiring specialized equipment and nursing staff [6]. In addition, various parameters have to be adjusted for each patient, e.g., duration of the process, the filtration rate and flow rates of the blood and dialysate. Clinical prediction models can aid in decision making by providing nephrologists with more accurate prognostic information under uncertainty of outcomes [7].

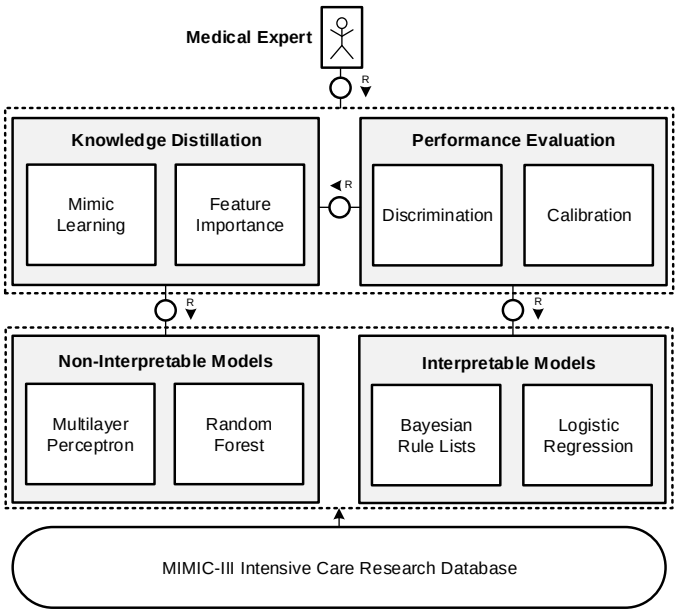


Figure 1. Our research setup modeled as a Fundamental Modeling Concepts (FMC) block diagram. Knowledge distillation approaches allow a medical expert to scrutinize the non-interpretable, black-box models.

In addition to usual criteria like accuracy or recall, when employing Machine Learning (ML) in the medical context, one especially important factor is the interpretability of the model, since doctors must take full responsibility for the respective decision, therefore requiring a high degree of trust [8]. As such, one can distinguish between two categories of ML algorithms: interpretable and non-interpretable. One example for interpretable models are Bayesian Rule Lists (BRL) [9]. By

presenting itself as *if...then...else* lists, it is easy for humans to comprehend both the decision making and the individual influence of each parameter on the outcome. In contrast, the Multilayer Perceptron (MLP) model is usually more accurate, but non-interpretable, since the weights of the nodes in the hidden layers are all that is exposed to the outside. Due to the fact that different loss and activation functions take effect when updating those weights, the abstraction to the original input data is too cumbersome for a human to grasp. By the same token, in the case of ensemble approaches such as Random Forest (RF), the number of constituent trees can be very high, e.g. $>$100, severely harming model intelligibility, even as the accuracy is improved.

In order to overcome the tradeoff between interpretability and accuracy, we employed knowledge distillation techniques, by means of which the complex inner workings of black-box algorithms are 'condensed' into easy-to-understand terms. Knowledge distillation is achieved, for example, by training an interpretable model on the predictions of a more accurate, non-interpretable model, a procedure termed mimic learning [10]. By means of this technique, we are able to gain insight into the complex model's decision process, thereby enhancing its intelligibility. As a further knowledge distillation technique, we utilized model-based feature importance for the RF model to visualize its most important features, illuminating the behavior the 'black box'.

Our contribution consists of developing and scrutinizing a Clinical Prediction Model (CPM) to prognosticate patient-specific outcomes after hemodialysis in the Intensive Care Unit (ICU). The research set-up is modeled in Figure 1 using a Fundamental Modeling Concepts (FMC) block diagram [11]. We evaluated the performance of two different model categories, BRL and Logistic Regression (LR) as the interpretable variants, along with MLP and RF as their non-interpretable counterparts. After that, we employed mimic learning and feature importance to help overcome the tradeoff between accuracy and interpretability and provide some insight into the decision parameters of the non-interpretable algorithms. We then interviewed an expert in the field of Nephrology to scrutinize the models thus developed.

The remainder of the work is structured as follows: In Section II we place our work in the context of extant research. We present our incorporated data and models in Section III and present results of our work in Section IV. We discuss our findings in Section V followed by the conclusion in Section VII.

II. Related Work

Our work is positioned at the intersection of ML and interpretability approaches in the context of predictive modeling. For this reason, in the following, we provide an overview of existing prognostic models applied in hemodialysis using both traditional and ML-based methods. Additionally, we outline selected interpretability methods with which knowledge distillation can be achieved.

A. Predictive Models for Hemodialysis Outcomes

When it comes specifically to predictive models for hemodialysis outcomes that employ logistic or Cox regression, a clear focus on prediction of mortality for chronic hemodialysis patients can be ascertained. For instance, a predictive model developed by Marks et al. for a cohort of chronic kidney disease patients (N=3,396) presented limited results in the prediction of 5-years mortality with Area Under the Receiver Operating Characteristic Curve (AUCROC)=0.753 [12]. For 60-day mortality of maintenance hemodialysis patients, Cohen et al. achieved AUCROC=0.87, albeit in a relative small cohort of 514 patients from eight clinics [7]. Finally, a systematic literature review and external validation study conducted by Ramspek et al. indicated that AUCROC of the models validated ranged from 0.710 to 0.752 with Floege et al.'s model being the best-performing, with AUCROC=0.79 in their original population (N=11,508) [14, 13].

ML research in Nephrology has been traditionally geared towards kidney disease detection using decision trees and naïve Bayes [15, 16]. However, those models tend to be less accurate when compared to more advanced models, which prompted the community to experiment with other methods, such as Support Vector Machines (SVM) and Artificial Neural Network (ANN) for prediction of kidney disease with encouraging results [17, 18]. In a similar fashion, Lakshmi et al. compared the three models, namely, logistic regression, random forest and ANN, proposing the latter for better performance and accuracy [19].

In the specific context of hemodialysis outcomes, ML approaches have also been employed, achieving some degree of success in the chronic setting. For example, Martínez-Martínez et al. employed a range of different ML methods, such as SVM and MLP, to predict hemoglobin levels and thereby anemia in a cohort of N=13,011 patients, achieving the lowest mean absolute error (0.662) with a bagging approach [20]. Furthermore, in a comparison of three different techniques, ANN, LR and Decision Tree (DT), Srisawat et al. recommended ANN for the mortality prediction task [21].

For critically ill patients, based on a cohort of N=76 Srisawat et al. found a panel of urinary biomarkers to be strongly predictive of renal recovery, presenting an AUCROC of 0.94. Regardless of the small sample size which demands more thorough validation, the needed biomarkers are not necessarily always available in an intensive care setting, potentially limiting the applicability of this biomarker panel.

B. Knowledge Distillation

The increasing complexity of ML models and the many parameters influencing their output make it considerably difficult – if not impossible – for a human to understand the influence of any specific feature on the training and outcomes of the model. Case in point are the weights of the multiple neurons in a MLP or the potentially hundreds of trees in a RF. To enable us to 'peek into the black box' we employed the concept of knowlegde distillation put forth by Che et al. utilizing mimic learning [10]. In addition, algorithms such as RF make it possible to derive feature importance based on specific criteria such as mean decrease in impurity. This method provides even further insight into the algorithm's inner workings.

In the context of ML models and results, Doshi-Velez defines interpretability as the ability to explain or to present in understandable terms to a human [22]. In contrast, Lipton

sees interpretability as a "non-monolithic concept" which encompasses a host of "distinct ideas" [23]. Expanding on these ideas, a fledging community of researchers, deemed Fairness, Accountability, Transparency (FAT) academics, emphasizes, amongst others, explainability as one of the core principles for accountable algorithms [24]. This principle establishes that algorithmic decisions should be intelligible to end-users in "non-technical terms". In the context of this paper, we define interpretability as a *property of machine learning algorithms and their outputs which allows scrutiny by medical experts*. Under scrutiny, we mean the ability of doctors to 1) easily ascertain the 'reasoning' behind an algorithm's decision, 2) identify the most important features for the output and 3) illuminate possible biases within the model.

In effect, the enhanced performance with modern ML tools, however, is achieved at the expense of model interpretability. The ability to explain and interpret decision is a key requirement in medical applications. In the context of ML, Lipton places particular focus on identifying decision boundaries and ascertaining the influence of specific feature for improved interpretability [23]. Approaches have been developed to achieve interpretability of black-box models, such as the classification vectors approach by Baehrens et al. and the Locally-Interpretable Model-agnostic Explanations (LIME) by Ribeiro et al. [25, 26]. In particular, Katuwal and Chen applied the LIME technique for achieving interpretability of random forests for predicting ICU mortality, achieving accuracies of 80 % [8]. Still in the medical domain, Hayn et al. quantified the influence of individual features on particular decisions made by a random forest in clinical modeling applications [27].

Unlike previous work, we focus specifically on the task of outcome prediction of hemodialysis patients in intensive care while comparing two types of models side-by-side, one interpretable (BRL and LR) and another non-interpretable (MLP and RF). For aiding the interpretability of the complex models, we made use of the mimic learning technique as proposed by Che et al. in lieu of the LIME method employed in extant research, because we aim to obtain a global understanding of the model's inner workings rather than explain individual instances of classification [8, 10]. Che et al. used Gradient Boosting Trees as mimic learning model while we applied Bayesian Ridge Regression (BRR) since their output more closely resembles logistic regression, a technique widely employed in medicine.

Given the extant literature on hemodialysis and knowledge distillation, one can ascertain a lack of works 1) using ML with a focus on critically ill patients, 2) covering different outcomes, not only mortality prediction and 3) scrutinizing model features by means of interpretability approaches. This paper addresses these research gaps.

III. Methods

In the following, we share details about the methods and data employed for the clinical models developed. We used *RapidMiner* [28], which allowed us to prepare data, develop and cross-validate first models. The final models were subsequently implemented with the *scikit-learn* library [29] in Python 2.7. The data we used were provided by the MIMIC-III dataset [30] stored in an in-memory database via an Open Database Connectivity (ODBC) interface [31]. To evaluate the models, we utilized discrimination as measured by Area Under the Receiver Operating Characteristic Curve (AUCROC) and calibration using Brier score and calibration plots. A metric routinely used in the medical context, Diagnostic Odds Ratio (DOR) was also provided in combination with precision, recall and sensitivity [32].

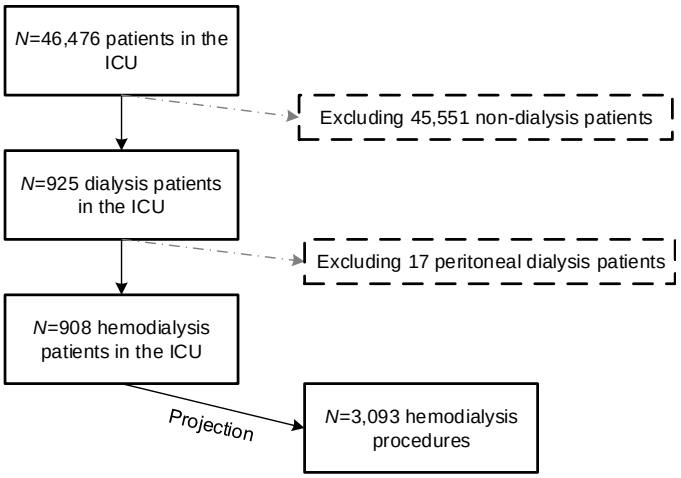


Figure 2. Cohort selection of the hemodialysis procedures based on the MIMIC-III intensive care patients.

A. MIMIC-III Database

The MIMIC-III intensive care research database contained hospital admission data for patients collected over an eleven-year period in a Boston hospital [30]. As seen in Figure 2, out of the approximately 46,000 patients present in the dataset, we extracted 908 relevant patients for this paper, totaling approximately 3,093 hemodialysis procedures for model training. We had to exclude from the analysis patients who had undergone peritoneal dialysis, who are not relevant in an acute context.

The cohort did not contain patients who underwent hemofiltration or hemodiafiltration, only hemodialysis patients. Under hemodialysis, the data comprises both Continuous Renal Replacement Therapy (CRRT) and Intermittent Hemodialysis (IHD) modalities, therefore RRT type was a feature in the final model. As such, we derived another cohort only with CRRT patients (N=1,163 procedures) and IHD patients (N=1,930 procedures) to ascertain whether results were consistent across hemodialysis modalities. We further derived a cohort consisting exclusively of acute patients (N=954 procedures), since patients who developed Acute Kidney Injury (AKI) without previous history of renal disease exhibit peculiarities from a clinical standpoint.

Missing Data: Due to the manually curated nature of the MIMIC-III dataset, aside from occasional data inconsistencies, a significant amount of data was missing. For example, the columns containing serum creatinine and Glomerular Filtration Rate (GFR) values before the procedure were missing in approx. 20 % of samples. As the scikit-learn models need a complete dataset for training, we decided to impute the missing values using k-nearest neighbors algorithm (k-NN) [33].

B. Features and Outcomes

In cooperation with a German university hospital, we conducted interviews in order to curate a list of suitable features, amounting to about 80 predictors. Those included patient demographics, such as age or Body-Mass Index (BMI), RRT parameters such as the duration of the procedure, comorbidities as well as laboratory values, including parameters such as serum creatinine and GFR for 24, 48 and 72 hours before the procedure and patient vitals.

Additionally, we included outcomes such as 90-day mortality, renal recovery, mechanical ventilation days and length of stay in the ICU. The variables ventilations days and length of stay presented continuous values, which had to be binarized for the BRL classifier to work, since it only supports binary outcomes. The complete list of features can be examined in Table A.I. The outcomes were thus defined:

- **90-days Mortality:** Indicates whether the patient has died within a 90-day period (1 = dead / 0 = alive),
- **Renal Recovery:** If patient has been for more than 7 days without hemodialysis requirement, renal function is considered to be restored (1 = recovery / 0 = no recovery),
- **Ventilation Days:** Indicates whether the patient has been on ventilation for been less than seven days (1 = true / 0 = false), and
- **Length of Stay:** Points out if length of stay has been less than 7 days (1 = true / 0 = false).

C. Modeling Algorithms

In the following, we describe the models and strategies used as well as the parameters chosen for training for both the interpretable and non-interpretable algorithms.

1) Bayesian Rule Lists: We chose the existing Python 2 implementation of BRL [9]. Letham et al. describe it as a direct competitor to decision tree approaches, as the model achieves high accuracy for classification tasks while still being intelligible for subject-matter experts. This algorithm tries to derive *if...then...else* statements over a dataset with the important criteria of their being sparse for better human readability. It builds Bayesian association rules consisting of an antecedent a and a consequent b. The consequent has a multinomial distribution over all the predicted labels y, so that the rules are defined by Equation (1):

$$a \rightarrow y \sim Multinomial(\theta) \tag{1}$$

The rules are generated by mining antecedents directly from the data and afterwards computing the posterior consequent distribution over the antecedent lists. BRL have the advantage of being easy to interpret due to their sparsity while retaining accuracy in classification. However, there are algorithms providing a higher accuracy, which also have the capability of more elaborate parameter tuning. Additionally, the current implementation of BRL has the shortcoming of a very long runtime and only being able to classify binary targets. Thus, we had to adjust the target features accordingly through use of a binary operator for continuous predictors.

Parameters: The sole adjustable parameter in the implementation used was the maximum number of iterations. Multiple adjustments to this parameter – including changes by a factor of ten – did not result in a significant change, neither for the runtime nor for the accuracy. For the evaluation, we chose a value of 50,000 maximum iterations.

2) Logistic Regression: LR is widely used for clinical prediction model development. It provides fast training time and easy-to-interpret coefficients for each model feature. For the sake of illustration, in a univariate logistic regression model, the probably that an input vector X can be assigned to the default class (or y =1, i.e., AKI onset) is given by Equation (2) also known as logit function:

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}} \tag{2}$$

The parameters β_0 and β_1 are not known and therefore must be estimated. This algorithm seeks to derive coefficients β_i for each input feature so that they map to a binary output while minimizing the error between predicted and actual class membership using maximum-likelihood estimation [34].

Owing to its simplicity, however, LR tends to perform worse when compared to more sophisticated algorithms such as MLP or RF. Critically, LR is built upon the assumption of linearly correlated inputs and outputs. This is potentially an issue, since in a medical context one cannot necessarily assume linear relationships.

Parameters: One of the key hyperparameters to be tuned for LR refers to the regularization strength. Model performance upon validation can be improved by penalizing large coefficients, potentially reducing overfitting. As such, model sparsity is improved by a strong regularization, typically defined as λ. Another key parameter to tune is the the type of penalty for the regularization, namely L1 (lasso) or L2 (ridge). Since utilizing L1 penalty shrinks the coefficients of less importance to zero, some features might be removed altogether, a desirable property when dealing with wide datasets. For our experiment, we chose $\lambda = 1$ and L1 regularization.

3) Multilayer Perceptron: We chose the scikit-learn implementation of MLP, which is able to handle both regression and classification tasks. Just as other implementations, this network consists of multiple layers of so-called "neurons": one input layer with as many neurons as there are inputs, one output layer with the size of the number of target features and hidden layers varying in size and quantity. The log-loss function is optimized through updating weights for each neuron for each iteration of model training. The neural network can be defined as mathematical function $f(x)$ as shown in Equation (3) with the activation function K and k-times $g_i(x)$ representing the dependencies between functions with an individual weight w_i.

$$f(x) = K\left(\sum_{i=1}^{k} w_i g_i(x)\right) \tag{3}$$

MLP is a widely used algorithm in ML due to its versatility and potentially high accuracy. It provides a wealth of parameters to tune. As such, finding the right ones for a specific use case can prove cumbersome. Furthermore, the decision making process

of such a neural network is not comprehensible to a human and thus provides nearly no interpretability.

Parameters: The amount of parameters to be adjusted when using neural networks is very extensive. Performing grid search over selected parameters, we found the default ones provided by the library to perform the best. This means the learning rate, which determines the speed and accuracy of convergence, was set to 0.001. The activation function, determining the output of the neurons in the hidden layer, was the rectifier linear unit "relu". The network consisted of one hidden layer with 100 neurons. We set the maximum number of iterations before convergence to 200.

4) Random Forest: The RF algorithm builds an ensemble of multiple trees in order to get a more accurate and stable prediction in comparison to an approach that relies on single decision tree. The ensemble's constituent trees utilize a random subset of the features available to split the nodes to be classified [35]. As a result of 'pooling' or majority voting of individual predictions, characteristically, RF are less prone to overfitting than regular decision trees. RF relies on bagging or bootstrap aggregation, i.e., sampling with replacement, to select samples of the training data, in an effort to reduce variance in the prediction function [36]. Hastie et al. formalize the concept in Algorithm 1.

Given a set of constituent trees b where $b \in \{1, \ldots, B\}$, we denote the overall class prediction of the random forest rf over all B trees for input x by $\widehat{C}^B_{rf}(x)$. Accordingly, if we denote the class prediction of the bth constituent tree by $\widehat{C}_b(x)$, the classification output of the RF model is given by Equation (4):

$$\widehat{C}^B_{rf}(x) = majority\ vote\{\widehat{C}_b(x)\}_1^B \tag{4}$$

Algorithm 1: Training a Random Forest

Input: Training Data
Result: Ensemble of Trees
for $b = 1$ *to* B **do**
- (a) Obtain bootstrap sample of size N from training data;
- (b) Grow tree T_b to the bootstrapped data, applying these steps recursively, until minimum node size n_{min} is reached:
  - i. Select m variables at random from the available p variables;
  - ii. Pick the best variable/split point among m;
  - iii. Split the node into two daughter nodes;

end
Return Ensemble of Trees $\{T_b\}_1^B$;

Parameters: The RF algorithm tends to perform well even without extensive tuning, what may explain its wide popularity [36]. In addition to the usual hyperparameters for decision trees, such as tree depth, the library employed exposes a number of hyperparameters that can be tuned specifically for RF. They include, e.g., the number of constituent trees (or estimators), i.e., B from Algorithm 1, number of variables m to split a node and the minimum number of leaves required to split an internal node. We determined the best hyperparameter combination for our use case via gridsearch, with a total number of estimators of 300, maximum tree depth of 16 and maximum number of features of eight.

D. Knowledge Distillation

In the following, the knowledge distillation approaches employed are presented in detail.

1) Mimic Learning: To provide some insight into the workings of the complex models employed we utilized a method called mimic learning. Building upon the approach of Che et al. we trained an interpretable model – the thus termed mimic model – on the outputs of the non-interpretable models, i.e., MLP and RF. In this approach, the mimic model takes on the same input features as the non-interpretable model.

In the case of MLP, the outputs of the non-interpretable model are termed soft scores. More generally, they are called prediction probabilities, meaning continuous variables approximating the actual prediction target. Training the mimic model on the prediction probabilities allows us to create a much smaller, thus understandable, faster but still comparably accurate model. In fact, under certain circumstances, it is even possible for the mimic model to generalize better than the non-interpretable model [10]. This happens because the non-interpretable model filters out certain noise in the training data, which could have a negative impact on training performance of the interpretable model. For the mimic model, we needed an algorithm which was able to predict continuous scores in order to train it on the aforementioned soft scores. For this purpose, we utilized BRR.

Similarly to common linear regression, BRR tries to find coefficients for each input feature so that they map to the target feature, minimizing loss. In addition to parameters common to linear regression, it includes regularization parameters to control the growth of the coefficients. Therefore, this model is less prone to over-fit while still being as fast as linear regression. Furthermore, regression in general has the advantage of being very fast concerning training time and interpretable, as one can easily inspect the coefficients for each feature. However, due to the simplicity of regression models, they usually lack accuracy when compared to more elaborate algorithms. Very few parameters can be adjusted for this algorithm and for our experiments, we applied the default ones. This means that all regularization parameters were set to 10^{-6} and the number of iterations before convergence was set to 300.

The process logic implemented for the mimic learning approach for MLP is shown in pseudo-code in Algorithm 2. A similar logic can be followed for RF, in which case the soft scores are replaced by prediction probabilities.

2) Feature Importance: Besides the aforementioned mimic learning approach, we provided feature importance metrics for RF the algorithm. In tree-based methods such as RF, one can estimate the relative feature importance by computing the decrease in node impurity by using it as split criterion. This decrease is averaged across all constituent trees and weighted proportionally to the number of samples it splits, i.e., nodes closer to the root of the tree will be deemed more important [37].

If we define $v(s_t)$ as the variable used in split s_t and $p(t) = N_t/N$ as the proportion of samples reaching t, the importance of a variable X_m over all N_T trees, i.e., $Imp(X_m)$, is defined by Equation (5). Note that $p(t)\Delta i(s_t, t)$ represents the weighted decrease in impurity over all nodes t which include X_m.

$$Imp(X_m) = \frac{1}{N_T} \sum_{T} \sum_{t \in T: v(s_t) = X_m} p(t)\Delta i(s_t, t) \quad (5)$$

Algorithm 2: Mimic Learning with BRR

Input: MLP Model, Training Dataset and Test Dataset
Result: Sorted mimic regression coefficients
Obtain soft scores from MPL on Training dataset;
Train BRR model on soft scores and Training dataset;
Apply trained BRR model on Test dataset;
Obtain BRR regression coefficients on Test dataset;
Sort regression coefficients;
Return Regression Coefficients;

IV. RESULTS

In the following section, we compare the performance of our interpretable models, BRL and LR, and our non-interpretable models, MLP and RF in terms of discrimination and calibration. Further, we present the knowledge distillation results of applying mimic learning to both MLP and RF and inspecting the feature importances of RF, since these were often the best-performing algorithms.

A. Model Performance

In the following, we assess model performance along three dimensions, discriminative power, calibration and computational performance in terms of runtimes.

1) Discrimination: Table I shows the overall performance of the employed classifiers according to the AUCROC performance metric. As expected, the MLP outperforms the BRL classifier in virtually every patient cohort and patient outcomes, excepting the prediction for ventilation days. The mimic approach using BRL trailed right along the MLP, presenting somewhat similar results. It worth noting that, in general, RF presented comparable performance to MLP, excepting the renal recovery task, in which MLP displayed more favorable results (0.91 vs. 0.83). In particular, the cohort of IHD patients presented similarly high AUCROC values for MLP, BRL, and BRR in the task of renal recovery ($\geq$ 0.9). This result suggests that patients in this cohort who do recover renal function possess very strongly discriminative features, which were captured by the algorithms.

While critically important, AUCROC is limited in the extent to which it can be used as sole metric to compare classifiers. Particularly in the medical domain, the trade-off between sensitivity (recall) and precision is highly dependent on the concrete use case. Therefore, we present further metrics in Table II for the outcome 90-days mortality in the complete patient cohort. This table shows that RF presented the best results across the metrics under analysis. Furthermore, while in terms of AUCROC MLP and RF do not differ substantially, the exception being the outcome renal recovery, there are marked differences when it comes to the other discrimination metrics, particularly DOR. LR presents overall poor results, displaying the lowest DOR. In combination its limited AUCROC, these metrics suggest that LR is likely an ill-suited choice for the task at hand in comparison with other modeling approaches. In effect, as illustrated by Table III, in comparison with previous discrimination results for similar albeit not identical tasks, our best model for renal recovery performed as well as the biomarker-based method proposed by Srisawat et al. [21].

```
IF SOFA: 0.69_to_inf THEN probability of DIED_90DAYS:
80.3% (73.1%-86.6%)

ELSE IF CR_24_B: 0.153_to_inf AND ELIXHAUSER: -inf_to_0.31
THEN probability of DIED_90DAYS: 3.0% (1.0%-6.1%)

ELSE IF LACTATE: 0.015_to_0.056 AND CR_72_B: -inf_to_0.18
THEN probability of DIED_90DAYS: 35.4% (29.5%-41.6%)

ELSE…
```

Figure 3. Excerpt of the rules from the Bayesian Rule Lists classifier when predicting 90-day mortality. Abbreviations: SOFA = Sequential Organ Failure Assessment score, CR_24_B, CR_72_B = Serum Creatinine 24h and 72h before procedure, respectively.

2) Calibration: Calibration estimates the agreement between predicted and observed risk. This is particularly relevant when it comes to predictive models employed in prognostic settings, such as ours, in which one is interested to predict future risk. In particular, it is possible for a model to be highly discriminative while over/underestimating risk, i.e., presenting poor calibration [38].

Figure 4 presents calibration curves for both MLP and RF for the outcome 90-days mortality in the complete patient cohort. In general, the MLP presents better calibration in comparison to RF. Nevertheless, we can observe that the MLP classifier tends to slightly underestimate the risk of death as the actual risk increases. In contrast, the RF classifier overestimates the probability of death in the low risk zone, with an inverse relation as the actual risk increases. We employed sigmoid and isotonic calibration to examine whether calibration could be improved. A slight improvement could be obtained for MLP, but in the case of MLP the calibration did not have the desired effect.

3) Runtimes: Concerning runtimes, there were considerable differences between the employed classifiers. While the MLP and RF took only a few seconds to conduct the full training with the configuration described previously, the BRL needed up to one hour to train on the same data. Due to the interpretable nature of the BRL, a medical expert can analyze the importance of single features directly on the model output.

B. Knowledge Distillation

Figure 3 shows the influence of some features and their values on the prediction of 90-days mortality for the complete cohort with the BRL algorithm. For this outcome, the Sequential Organ Failure Assessment (SOFA) score was a key feature. This score is widely used in intensive care for this very purpose, therefore the BRL classifier correctly detected

TABLE I. Simulation results displaying AUCROC for the different analysis cohorts and patient outcomes. Abbreviations: IHD = Intermittent Hemodialysis, CRRT = Continuous Renal Replacement Therapy, MLP = Multilayer Perceptron, RF = Random Forest, BRL = Bayesian Rule Lists, LR = Logistic Regression and BRR = Bayesian Ridge Regression, LOS ICU = Length of Stay in the ICU.

Outcome	Complete cohort					Acute patients					IHD patients					CRRT patients				
	MLP	RF	BRL	LR	BRR	MLP	RF	BRL	LR	BRR	MLP	RF	BRL	LR	BRR	MLP	RF	BRL	LR	BRR
90-days mortality	**0.84**	**0.84**	0.76	0.71	0.79	0.83	**0.85**	0.79	0.79	0.81	**0.83**	0.82	0.79	0.69	0.79	0.77	**0.78**	0.72	0.66	0.72
Renal Recovery	**0.91**	0.83	0.88	0.77	0.88	**0.86**	0.83	0.68	0.72	0.79	**0.92**	0.81	0.90	0.76	0.90	**0.86**	0.73	0.79	0.70	0.84
Ventilation Days	**0.81**	0.79	0.75	0.74	0.80	0.64	0.64	**0.68**	**0.68**	0.65	**0.81**	0.74	0.78	0.73	0.79	0.77	0.64	**0.79**	0.64	**0.79**
LOS ICU	**0.83**	0.80	0.82	0.73	0.82	**0.78**	0.64	0.69	0.63	0.73	0.80	**0.82**	0.78	0.78	0.80	0.73	0.64	**0.73**	0.63	**0.73**

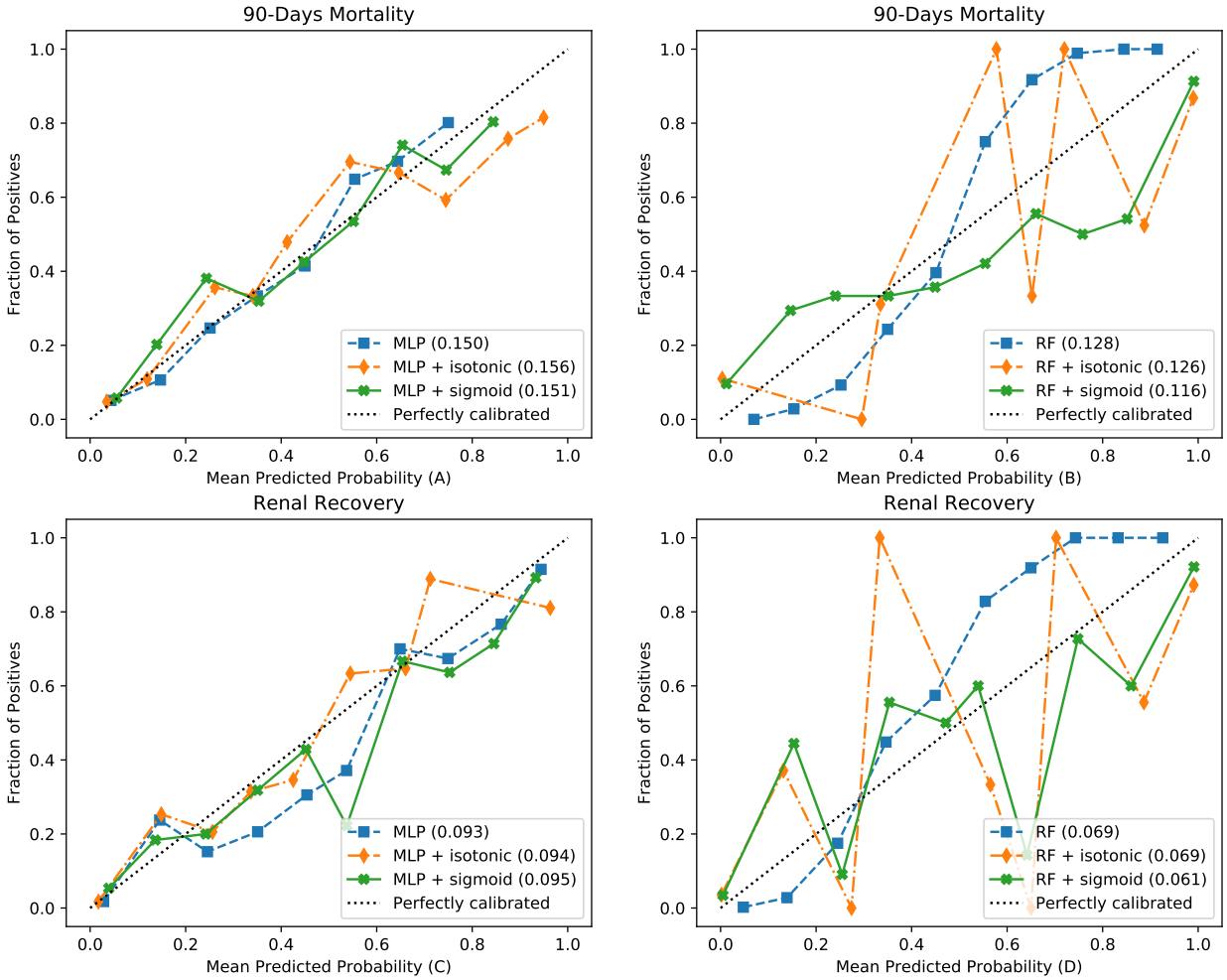


Figure 4. Model calibration depicted for Multilayer Perceptron (MLP) and Random Forest (RF) for the outcomes of 90-days mortality and renal recovery.

TABLE II. Overview of key discrimination metrics for the outcome 90-days mortality in the complete patient cohort. Abbreviations: MLP=Multilayer Perceptron, RF=Random Forest, BRL=Bayesian Rule List, LR=Logistic Regression, DOR=Diagnostic Odds Ratio.

Algorithm	Precision	Recall	Specificity	DOR
MLP	0.76	0.73	0.83	13.84
RF	0.88	0.77	0.92	41.45
BRL	0.76	0.64	0.85	10.5
LR	0.68	0.66	0.77	6.73

TABLE III. Overview of CPMs for used in the context of hemodialysis outcomes. Abbreviations: CPM=Clinical Prediction Model; N=number of patients; AUC=Area Under the Curve

CPM	*N*	End point	AUC
Marks et al.	3,396	Mortality	0.75
Cohen et al.	514	Mortality	0.87
Floege et al.	11,508	Mortality	0.79
Srisawat et al.	54	Renal recovery	0.94
Our approach	908	Renal recovery	0.92

this. "CR_24_B" corresponds to blood creatinine 24h before the hemodialysis procedure and Elixauser is a comorbidity score. High values for both of these features are associated with increased mortality, but from the output of the BRL alone it is hard to ascertain whether it correctly captured this relationship.

For the MLP and RF results to be inspected, we had to

apply the mimic learning strategy discussed previously. First, we needed to evaluate if the performance of the mimic model is satisfactory when being trained on the outputs (soft scores) of the MLP. One can verify in Table I that, while the BRR is still worse than the MLP as a rule, it performed better than the BRL, even if by a small margin. It is important to highlight, however, that the mimic classifier is only as good as the predictor it originally learned from.

In Figure 5, we can assess the influence of single features on a positive prediction of both 90-day mortality and recovery of renal function. The chart depicts the regression coefficients of the BRR mimic model for MLP and RF. The sign of the coefficients determine the direction of correlation and the absolute component represents the magnitude. For example, the higher the rightmost feature, e.g., the age of the patient, the higher is the probability of the patient to die within 90 days. Conversely, the higher the features with negative coefficients, e.g., the hemoglobin value in the blood of the patient, the less likely the patient is to die within 90 days. These results were submitted to the appraisal of a Nephrology expert to establish clinical relevance and adequacy.

In addition to the mimic learning results, the RF makes it possible to derive feature importances, which enables non-ML experts to have a sense of how individual features contribute to the outcomes. Figure 6 displays feature importances for 90-days mortality in the complete cohort, reverse-ordered by feature importance. Note that we show only the top 20 features in Figure 6. Unlike the coefficients of the mimic learning approach, the information regarding the direction of correlation is not readily available, solely the magnitude of importance.

V. Discussion

In the following, we will examine the model performance in light of related work and the insights obtained via knowledge distillation.

A. Model Performance

From a classification performance standpoint, our experiments suggest MLP and RF as suitable classifiers for the given prediction tasks, with BRL as a close third. In fact, both MLP and RF performed particularly well for renal recovery prediction, a key outcome for nephrologists. In effect, the positive discriminative results obtained with these two classifiers are consistent with previous work targeted at other but similar tasks [19, 17]. When it comes to dialysis outcomes models, based on discriminative performance as measured by AUCROC, our models outperform existing work based on logistic regression approaches, except for the mortality prediction outcomes of Cohen et al. [12, 7, 14]. These works are based on the chronic setting, though. A direct comparison of the models' performance, while not advisable, gives us at least a benchmark against which to compare, since the works in the acute setting are lacking in the literature.

The only work considering a cohort of acute patients, the biomarker-based approach by Srisawat et al., performs better than our model in the task of renal recovery, but the small sample utilized in the study might compromise its generalizability [21]. Furthermore, the needed urinary biomarkers might not be readily available at all times in the ICU setting. Potential cost considerations for these biomarkers should also be taken into account.

In spite of the promising results, upon closer examination, the approaches we developed have issues that might hinder their adoption in clinical practice. If we examine, e.g., the overall best-performing classifier, RF, for the outcome 90-days mortality, it presents a higher specificity than sensitivity (recall), meaning that it will fail to acknowledge high-risk patients (true positives) more often than it identifies low-risk patients (true negatives). This behavior of the RF classifier is further illustrated by the calibration curves in Figure 4. The implications of this difference must be examined in the context of the specific clinical use case and can be mitigated with careful threshold selection and other calibration techniques.

B. Knowledge Distillation

When it comes to ML models deployed in sensitive domains, discriminative performance is not enough. The models must be scrutinized with regards to their medical relevance and physiological meaningfulness. For example, some of the features deemed important for the MLP classifier do make sense from a medical standpoint, such as higher age correlating with a higher chance of mortality. However, the results also indicate that high levels of Glomerular Filtration Rate (GFR), a measure of how well the kidneys are functioning, is associated with higher mortality, a counterintuitive outcome, since physiologically it represents a protective factor. The mimic model for the RF classifier captures similar variables as the ones found in the mimic MLP, albeit with different coefficients, i.e., with differences in magnitude. However, the same criticism can be leveled at it: GFR features prominently in it as a risk factor instead of protective factor.

In a similar fashion, for the renal recovery outcome, both mimic MLP and mimic RF captured similar features. In this case, the positive outcome (recovery) is one (1) and the negative outcomes (no recovery) is zero (0). High hemodialysis dosage, therefore, would correlate with a higher likelihood of recovery. However, there is considerable debate in the medical literature as to whether higher dosage leads to better outcomes [12]. Therefore, this result must be interpreted with caution. Furthermore, the Sequential Organ Failure Assessment (SOFA) score appears to be a factor *favoring* renal recovery in the case of mimic RF, what clearly contradicts clinical expectations. A hypothetical explanation for this scenario is that patients with high SOFA scores are particularly ill and therefore receive, on average, better standard of care than other less severely ill patients. This hypothesis suggests that the prediction results should be further stratified by disease severity.

It is important to note that these potential spurious correlations are only illuminated through model interpretability, be it because of the nature of the model or the application of mimic learning. Thus, the model interpretability approach employed gives us the possibility to examine the correlations and question assumptions which otherwise might just go unnoticed when using non-interpretable models. However, usually there are non-linear correlations between certain blood values and outcome (e.g., U-shaped curve), such as potassium, as either too low or too large values can influence the patient's

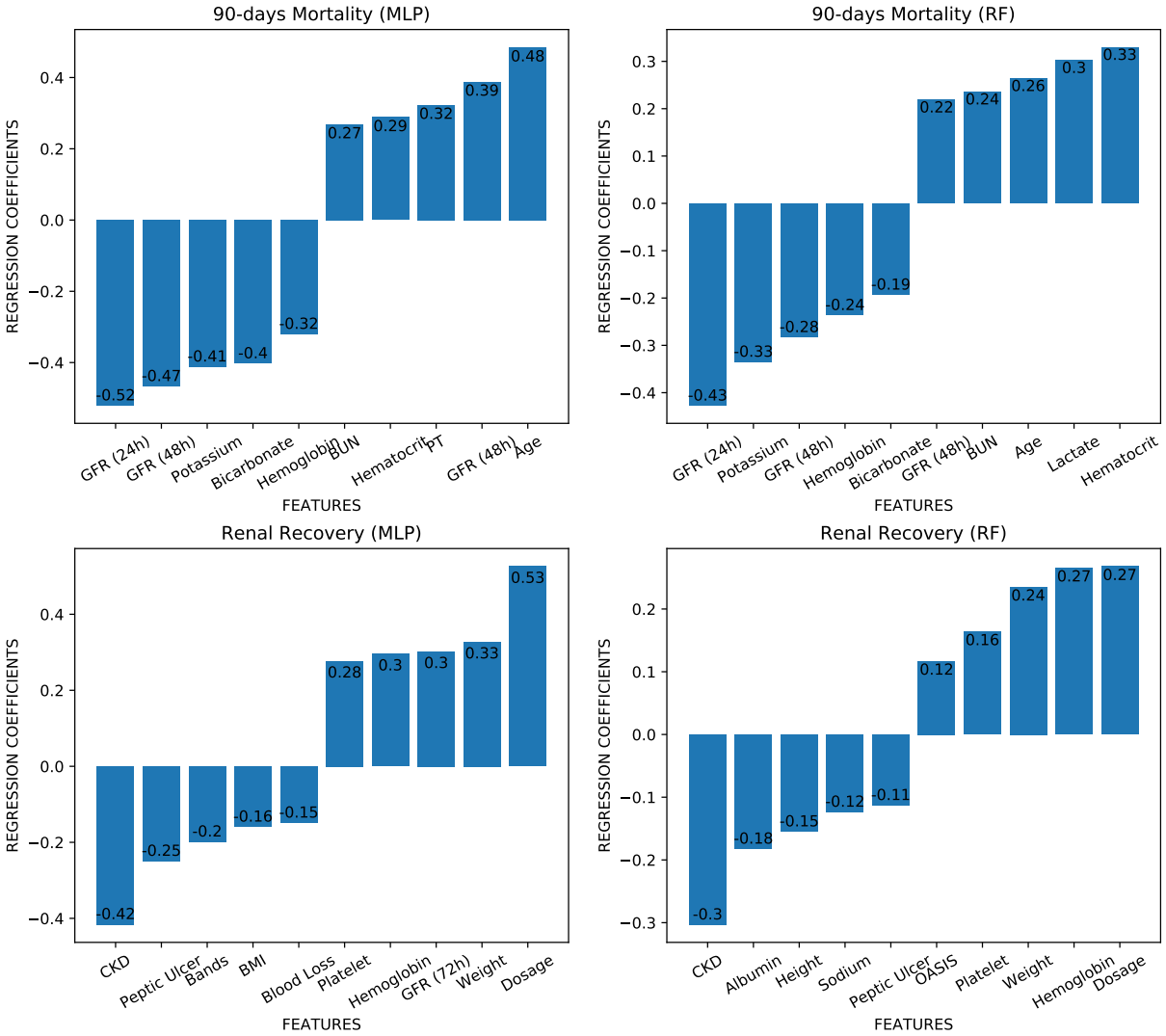


Figure 5. Coefficients of the most important features for the Bayesian Ridge Regression trained as mimic model for 90-days mortality and renal recovery for both Multilayer Perceptron (MLP) and Random Forest (RF). Abbreviations: GFR = Glomerular Filtration Rate 24h and 48h before procedure, respectively; BUN = Blood Urea Nitrogen; CKD = Chronic Kidney Disease; OASIS = Oxford Severity of Illness Score.

health negatively. Such relationships cannot be adequately represented by the mimic learning approach utilized.

To a certain extent, the feature importances of the RF classifier reflected the knowledge gleaned from applying the mimic learning approach, highlighting some of the parameters that also were captured in the mimic learning method, such as hematocrits, lactate, blood urea nitrogen and glomerular filtration rate. Despite this fact, when examining all approaches in combination, there is disagreement, for instance, in the magnitude of contribution or in how often a feature is mentioned across different techniques. Hall and Gill recommend that researchers combine different interpretability approaches in order to obtain a more intelligible picture of the model's behavior [39].

Finally, algorithms considered to be interpretable might not necessarily be intelligible. This is particularly evident for the BRL algorithm. Take its output as depicted in Figure 3. As a matter of fact, higher lactate values usually lead to other complications, but the upper bound of "infinity" is not meaningful in clinical practice. In order to refine and validate those assumptions, it is necessary to further analyze the data. Finding actual upper and lower bounds in the dataset can provide some insight into the actual values the model considers when making predictions.

VI. Limitations

Even though we achieved satisfactory discriminative performance, this analysis was based on a comparatively small patient cohort (N=908). Therefore, a validation study with a larger cohort is needed in order to derive generalizable claims. Additionally, missing data may have a significant influence on the quality of the predictions and certain features could be dropped if they are missing a large amount of values. We sought to mitigate this effect by means of multiple imputation with k-NN, but we cannot guarantee that no biases resulted from this approach.

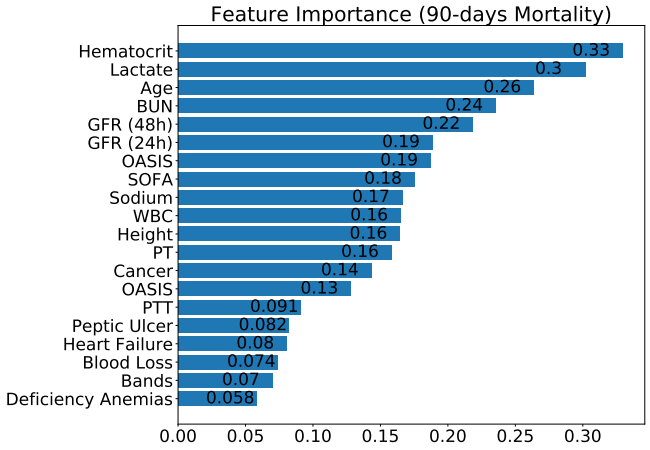


Figure 6. Feature importances for 90-days mortality based on the Random Forest (RF) classifier. Abbreviations: BUN = Blood Urea Nitrogen; GFR = Glomerular Filtration Rate 24h and 48h before procedure, respectively; OASIS = Oxford Severity of Illness Score; SOFA = Sequential Organ Failure Assessment; WBC = Whole Blood Count; PTT = Partial Thromboplastin Time.

Furthermore, in most cases, the mimic learning approach is fundamentally limited by the performance of the original model. In our experiments, the BRR performed worse when being trained on the outputs of the MLP as opposed to being trained directly on the real targets, because it most probably also assimilates the errors of the MLP. This can be ameliorated by improving the performance of the MLP through further parameter tuning.

Besides, the literature of interpretability approaches is growing. In this paper, we were able only to utilize two of them, mimic learning and method-based feature importances, thus necessarily providing a limited picture of model behavior. Other methods could be explored, for example involving local interpretability, such as Local Interpretable Model-agnostic Explanations (LIME) [26]. Finally, the medical relevance and physiological meaningfulness of the mimic models was evaluated by one expert only. Ideally, these should be assessed by a panel of expert to reduce biases.

VII. Conclusion

In this paper, we compared the performance of different models used in the prediction of hemodialysis outcomes, namely, 90-days mortality, ventilation days, length of ICU stay and renal recovery using data routinely acquired in a intensive care setting. The algorithms employed consisted of a combination interpretable and non-interpretable models. Our results suggest that ML approaches such as MLP and RF present satisfactory discriminative results (AUCROC $\leq$ 10 in the case of renal recovery) when compared with interpretable algorithms, such as LR or BRL.

However, an important aspect is the interpretability of such models if they are to be used for decision support in a clinical setting. To this end, we applied a knowledge distillation technique called mimic learning along with feature importances in order to scrutinize the 'black-box' best-performing models. The use of these techniques made it possible to uncover potentially spurious correlations captured by the algorithms, therefore shedding light on model biases. Therefore, we urge researchers who rely on ML for clinical predictive modeling to include an assessment of possible biases using for example knowledge distillation approaches.

Future work could take the form of further data analysis and processing, i.e., inclusion of more features, more elaborate imputation strategy and collection of more information about the patients. Besides, deployment in a clinical setting requires external validation using datasets from different institutions. Subsequently, an impact analysis of the use of such models in a clinical setting should be conducted to ascertain the impacts on care delivery and patient outcomes.

Acknowledgments

Parts of the given work were generously supported by the European Union's Horizon 2020 research and innovation program under grant agreement No 780495.

References

[1] H. F. da Cruz, S. Horschig, and M. Schapranow, "Prediction of Patient Outcomes After Renal Replacement Therapy in Intensive Care," in *Proceedings of the 3rd Intl. Conf. on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing (HEALTHINFO)*, vol. 3, 2018, pp. 7–12.

[2] E. Marieb and S. M. Keller, *Essentials of Human Anatomy & Physiology*. Pearson, 2018.

[3] C. Ronco *et al.*, "Renal Replacement Therapy in Acute Kidney Injury: Controversy and Consensus," *Critical Care*, vol. 19, no. 1, pp. 1–11, 2015.

[4] G. M. Fleming, "Renal Replacement Therapy Review," *Organogenesis*, vol. 7, no. 1, pp. 2–12, 2011.

[5] S. Negi *et al.*, "Renal Replacement Therapy for Acute Kidney Injury," *Renal Replacement Therapy*, vol. 2, no. 1, p. 31, 2016.

[6] L. Forni and P. Hilton, "Continuous Hemofiltration in the Treatment of Acute Renal Failure," *New England Journal of Medicine*, vol. 336, no. 18, pp. 1303–1309, 1997.

[7] L. M. Cohen *et al.*, "Predicting Six-month Mortality for Patients Who Are on Maintenance Hemodialysis," *Clin J Am Soc Nephrol*, vol. 5, no. 1, pp. 72–79, jan 2010.

[8] G. J. Katuwal and R. Chen, "Machine Learning Model Interpretability for Precision Medicine," https://arxiv.org/abs/1610.09045 [retrieved: August, 2018], Oct 2016.

[9] B. Letham *et al.*, "Interpretable Classifiers Using Rules and Bayesian Analysis: Building a Better Stroke Prediction Model," *Annals of Applied Statistics*, vol. 9, no. 3, pp. 1350–1371, 2015.

[10] Z. Che *et al.*, "Interpretable Deep Models for ICU Outcome Prediction," in *Proceedings of the AMIA Annual Symposium*, vol. 2, 2016, pp. 371–380.

[11] A. Knöpfel, B. Gröne, and P. Tabeling, "Fundamental modeling concepts," *Effective Communication of IT Systems, England*, p. 51, 2005.

[12] A. Marks *et al.*, "Predicting Renal Replacement Outcomes in a Large Community Cohort with Chronic Kidney Disease," *Nephrol Dial Transplant*, vol. 30, no. May, pp. 1507–1517, 2015.

[13] C. L. Ramspek *et al.*, "Prediction Models for the Mortality Risk in Chronic Dialysis Patients: a Systematic Review and Independent External Validation Study," *Clinical Epidemiology*, vol. 9, pp. 451–464, 2017.

[14] J. Floege *et al.*, "Development and Validation of a Predictive Mortality Risk Score from a European Hemodialysis Cohort," *Kidney International*, vol. 87, no. 5, pp. 996–1008, 2015.

[15] P. S. Baby and P. Vital, "Statistical Analysis and Predicting Kidney Diseases using Machine Learning Algorithms," *J of Eng Res & Tech*, vol. 4, no. 07, pp. 206–210, 2015.
[16] R. Greco *et al.*, "Decisional Trees in Renal Transplant Follow-up," *Transplant Proc*, vol. 42, no. 4, pp. 1134–1136, may 2010.
[17] S. Vijayarani and S. Dhayanand, "Kidney Disease Prediction Using SVM and ANN," *Comp Bus Res*, vol. 6, no. 2, pp. 6–17, 2015.
[18] P. Sinha and P. Sinha, "Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM," *J of Eng Res & Tech*, vol. 4, no. 12, pp. 608–612, 2015.
[19] K. R. Lakshmi, Y. Nagesh, and M. Veerakrishna, "Performance Comparison of Three Data Mining Techniques for Predicting Kidney Dialysis Survivability," *J of Adv in Eng & Tech*, vol. 7, no. 1, pp. 242–254, 2014.
[20] J. M. Martínez-Martínez *et al.*, "Prediction of the Hemoglobin Level in Hemodialysis Patients using Machine Learning Techniques," *Computer Methods and Programs in Biomedicine*, vol. 117, no. 2, pp. 208–217, 2014.
[21] N. Srisawat *et al.*, "Urinary Biomarkers and Renal Recovery in Critically," *Clin J Am Soc Nephrol*, vol. 6, pp. 1815–1823, 2011.
[22] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," no. Ml, pp. 1–13, 2017.
[23] Z. C. Lipton, "The Mythos of Model Interpretability," in *Proceedings of the Workshop on Human Interpretability in Machine Learning*, 2016, pp. 96–100.
[24] D. Shin and Y. J. Park, "Role of fairness, accountability, and transparency in algorithmic affordance," *Computers in Human Behavior*, vol. 98, pp. 277–284, 2019.
[25] D. Baehrens *et al.*, "How to Explain Individual Classification Decisions," *Mach Learning Res*, vol. 11, pp. 1803–1831, 2010.
[26] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," in *Proceedings of the Workshop on Human Interpretability in Machine Learning*, 2016, pp. 91–95.
[27] D. Hayn *et al.*, "Plausibility of Individual Decisions from Random Forests in Clinical Predictive Modelling Applications," *Studies in Health Technoly and Informatics*, pp. 328–335, 2017.
[28] M. Hofmann and R. Klinkenberg, *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. CRC Press, 2013.
[29] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J of Machine Learning Res*, vol. 12, pp. 2825–2830, 2011.
[30] A. E. W. Johnson *et al.*, "MIMIC-III, a Freely Accessible Critical Care Database," *Scientific Data*, vol. 3, 2016.
[31] F. Färber *et al.*, "SAP HANA database," *ACM SIGMOD Record*, vol. 40, no. 4, pp. 45–51, 2012.
[32] A. S. Glas *et al.*, "The Diagnostic Odds Ratio: A Single Indicator of Test Performance," *Journal of Clinical Epidemiology*, vol. 56, no. 11, pp. 1129–1135, 2003.
[33] J. M. Keller, M. R. Gray, and J. A. Givens, "A Fuzzy K-nearest Neighbor Algorithm," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 4, pp. 580–585, jul 1985.
[34] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. John Wiley & Sons, 2013, vol. 398.
[35] A. Gron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, 1st ed. O'Reilly Media, Inc., 2017.
[36] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009, vol. 1.
[37] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding Variable Importances in Forests of Randomized Trees," *Neural Information Processing Systems*, pp. 1–9, 2013.
[38] L. E. Hodgson *et al.*, "Systematic Review of Prognostic Prediction Models for Acute Kidney Injury (AKI) in General Hospital Populations," *BMJ Open*, vol. 7, no. 9, pp. 1–10, 2017.
[39] P. Hall and N. Gill, *An Introduction to Machine Learning Interpretability*, 2018.

APPENDIX

In Table A.I, we share the complete list of features used for our models.

TABLE A.I. Model features. Note that related features are grouped together. Abbreviations: Body-Mass Index (BMI), Acute Kidney Injury (AKI), Sequential Organ Failure Assessment (SOFA), Simplified Acute Physiology Score (SAPS), Partial Thromboplastin Time (PTT), International Normalized Ratio (INR), Prothrombin Time (PT), Whole Blood Count (WBC).

Category	Feature
Demographics	Age
	Height, Weight, BMI
	Ethnicity
	Gender
Hemodialysis-related	Dosage
	Modality
	AKI stage
Comorbidities	AIDS
	Alcohol abuse
	Blood loss anemia
	Cardiac arrhythmias
	Chronic pulmonary
	Coagulopathy
	Congestive heart failure
	Deficiency anemias
	Depression
	Diabetes complicated, Diabetes uncomplicated
	Drug abuse
	Elixhauser Vanwalraven score
	Fluid electrolyte imbalance
	Hypertension
	Hypothyroidism
	Liver disease
	Lymphoma, Metastatic cancer, Solid tumor
	Obesity
	Other neurological disorders
	Paralysis
	Peptic ulcer
	Peripheral vascular
	Psychoses
	Pulmonary circulation
	Renal failure
	Rheumatoid arthritis
	Valvular disease
	Weight loss
ICU scores	OASIS
	SOFA
	SOFA Renal
	SAPS
Vitals	Heart rate
	Systolic Blood pressure
	Diastolic Blood pressure
	Mean Blood pressure
	Respiratory Rate
	Temperature °C
	Oxigen Saturation (SpO_2)
Laboratory values	Anion gap
	Albumin
	Bands
	Bicarbonate
	Bilirubin
	Blood urea nitrogen
	Creatinine 24, 48 and 72h before procedure
	Chloride
	Glucose
	Hematocrit
	Hemoglobin
	Lactate
	Platelet
	Potassium
	PTT, INR, PT
	Sodium
	WBC
	Glomerular Filtration Rate 24, 48 and 72h before procedure

Identification of Appropriate Facial Expressions for Medical Doctors

Akemi Oe, Yoshitoshi Murata
Graduate School of Software and Information Science
Iwate Prefectural University
Iwate, Japan
e-mails: akemioe@gmail.com, y-murata@iwate-pu.ac.jp

***Abstract*— Establishing trust between a patient and a doctor depends as much on their relationship as on the doctor's medical abilities. One of important factors in building and maintaining a relationship is whether the doctor produces facial expressions appropriate to the patient's condition. The purpose of this study is to identify facial expressions appropriate to various patient conditions. We focused on the greetings given by young doctors at the beginning of medical interviews of adult patients in the general ward of a hospital. Images of a role-playing patient portraying one of three physical conditions were shown to seven student doctors, who were then videotaped as they greeted the "patient." We identified appropriate facial expressions for each condition on the basis of both human evaluation and computer-aided facial expression emotion analysis. We also identified an appropriate facial expression when a young doctor auscultates a patient and clarified differences in the evaluation of facial expressions due to the gender and age of the evaluators. Finally, we clarified facial expressions required for pediatricians.**

Keywords-doctor-patient interaction; facial expression; nonverbal communication.

I. INTRODUCTION

We study facial expressions required for young medical doctors, and presented research results in the international conference of Global Health 2018 [1]. In this paper, we introduce differences due to the gender and age of the patients and facial expression required for pediatricians in addition to the paper presented in the Global Health 2018.

Patient satisfaction is one of important components of medical care [2]. Improving patient satisfaction enhances trust and the relationship between patient and doctor, which leads to stronger adherence to the prescribed protocol, such as taking medicine, and to enhanced therapeutic effect [3][4]. Many studies and reviews have shown that the main determinant of patient satisfaction is the doctor-patient relationship [5]–[9] and that patient satisfaction is higher when the patient communicates with a doctor having strong nonverbal communication ability [10][11]. Unfortunately, inexperienced young doctors and medical students often have trouble producing appropriate facial expressions when greeting a patient. The first author of this paper, a lecturer on medical communication, often hears young doctors complaining that, though they intend to smile, patients say that they seem to be angry.

In this study, we identified appropriate facial expressions for a doctor by using quantitative analysis. We videotaped medical students' facial expressions when they greeted a patient and analyzed the recorded videos using computer-aided facial expression emotion analysis. We then asked potential patients to evaluate the appropriateness of the facial expressions on a five-point scale. We then identified facial expressions appropriate for doctors for three patient conditions.

The facial expression required for a doctor may be influenced by the patient's condition, the environment in which treatment is provided, culture, the medical department, and both patient's and doctor's gender. As the first step, we focused on appropriate facial expressions for young medical doctors when greeting patients in the general ward of a hospital. Also, as a typical medical treatment situation, we focused on appropriate facial expressions when young medical doctors auscultate a patient. We clarified the difference in the evaluation on facial expressions due to gender differences and age differences among evaluators. Finally, using computer-aided facial expression emotion analysis, we analyzed the facial expressions produced by two pediatricians when shown photographs of pediatric patients in three different conditions.

After reviewing related work in Section II, we explain our facial expression analysis system in Section III. We describe our evaluation experiment in Section IV and analyze facial expressions in recorded videos in Section V. Gender and age-specific differences among evaluators are discussed in Section VI, and appropriate facial expressions for pediatricians are presented in Section VII. We conclude with a summary of the key points in Section VIII.

II. RELATED WORK

The medical interview is important for a doctor not only to get relevant information from the patient but also to build a good relationship with the patient. Therefore, we review existing research related to medical interviews in this section. In addition, since our work focuses on facial expressions, we also review research on nonverbal communication.

A. Medical Interviews

Medical interviews have traditionally focused on gathering relevant information from patients [12]. Nowadays, the focus has expanded to building a trusting relationship, sharing decision-making, responding to the patient's

emotional state, and supporting actions related to the patient's condition and treatment, so the doctor must have a wider range of communication skills [13]. These skills include "looking at a patient not as a case but as a human being" [14] and "building and maintaining a good relationship between doctor and patient" [15]. It has been shown that such skills have a greater effect on patient satisfaction than the doctor's medical skills, the medicine prescribed, the information provided, the questions asked, the advice given, and the instructions given. In particular, a patient's satisfaction is positively related to the doctor being warm [14][16], empathic [14][16]–[18], and friendly [16] and giving the impression of being human [17].

"Nonverbal communication" is a means of communicating these emotional aspects. Patient satisfaction is higher when the doctor has a strong ability to express his or her emotions and to read the emotions of others by nonverbal communication such as through facial expressions, gaze, posture, and tone of voice [10][19][20]. In short, a doctor's nonverbal communication is an important aspect of patient care.

B. *Nonverbal Communication*

Facial expression plays a large role in nonverbal communication. For emotional messages such as "like" and "dislike," Mehrabian [21] estimated that 7% of the message is carried by the language content, 38% is carried by the voice and sound quality, and 55% is carried by the facial expression and gestures. Birdwhistell [22] argued that 35% of the message is carried by the language content while the remaining 65% is carried by the expression, the way of talking, the gesturing, etc. Therefore, it is useful to clarify the appropriate facial expressions for doctors to have when communicating with patients, especially patients who are sensitive to a doctor's nonverbal behavior due to anxiety [23] [24].

Differences in non-verbal expressions due to cultural background have also been pointed out. North American countries such as the United States and Canada and Western European countries such as the Netherlands, Italy, and Belgium are individualistic cultures, whereas Japan is considered to be a collectivist culture located at the opposite extreme. People living in collectivist cultures tend to be cautious about nonverbal expression [25]. Comparative cultural research on facial expression recognition in the UK, Italy, and Japan revealed no significant difference in cognitive ability between Japan and the other two countries while the expressions produced by Japanese tended to be ambiguous and difficult to recognize [26]. Other research has shown that women have better facial expression cognitive abilities than men [27]–[31].

Studies on the effects of age on facial expression recognition have focused on infants and disabled children, but recent studies have focused on elderly people as well. These studies have shown that elderly people consistently have trouble recognizing "anger," "sadness," and "fear" while their recognition abilities for "happiness," "surprise," and "disgust" do not exhibit a consistent tendency [32][33].

A comparison study between young and elderly people revealed that young people have a higher propensity to evaluate photographs of expressionless faces as showing "anger" than elderly people [34].

III. FACIAL EXPRESSION ANALYSIS SYSTEM

To identify appropriate facial expressions, we developed a system that quantitatively analyzes changes in facial expression. It is based on the Cognitive Services Emotion API [35] provided by Microsoft's Azure cloud service. The concept of the system is illustrated in Figure 1. The doctor's facial image during a patient interview is recorded in a video file. The file is sent to the Cognitive Services Emotion API, which provides feedback on the position coordinates of the doctor's face and the ratio by emotion. The system comprises Microsoft.ProjectOxford.Emotion.dll which corresponds to the Emotion API on Azure, Newtonsoft.Json.dll which handles the file in JSON format, Parakeet.dll which processes the movie file, and Parakeet.Logging.dll which passes the facial expression analysis received from the Emotion API as a log file to a PC for real-time display. We developed Parakeet.dll, Parakeet.Logging.dll, and LogViewerWPF.exe.

Our facial expression emotion analysis system calculates the ratio for seven emotions ("happiness," "anger," "contempt," "disgust," "fear," "sadness," and "surprise") reflected in the input video image and for "neutral." The total for all emotions is 1, and the value for "neutral" is obtained by subtracting the total value for the seven emotions from 1.

The detection result window is shown in Figure 2. Each row shows the results for the emotion corresponding to one facial expression. The rows are in time series order, with the latest set of results in the bottom row. Clicking the display command on the menu highlights the detection results for the specified face. Selecting a line by using the mouse or keyboard causes the corresponding video to play.

The cells corresponding to each emotional value are highlighted in "pink." If the value = 1, the cells are the darkest pink, and if the value = 0, the cells have no color. For example, if the emotional value is 0.8, the color density is 80%.

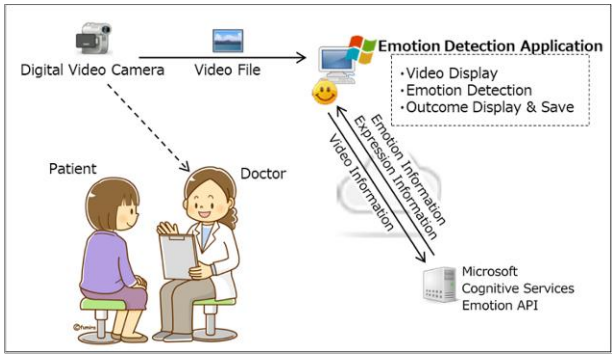


Figure 1. Configuration of video image-based emotion analysis system.

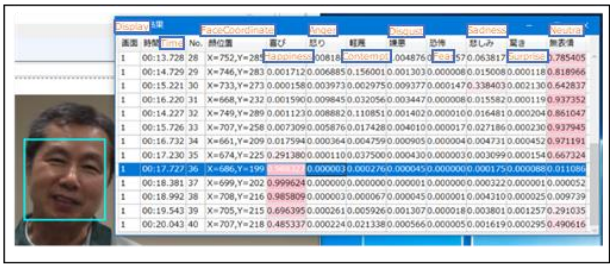


Figure 2. Detection result window.

IV. EVALUATION EXPERIMENT

One way to analyze the facial expressions of veteran doctors would be to identify facial expressions appropriate for doctors. However, their facial expressions would not always be appropriate. Since many young physicians have trouble producing appropriate facial expressions when greeting a patient, we chose to identify facial expressions that would be acceptable for most patients including potential patients from facial expressions thought to be suitable by medical students. We targeted the situation, in which they greet patients in the general ward of a hospital. We videotaped their greeting for model patients to be evaluated and identified appropriate facial expressions. Although evaluation by actual patients is best, it would have been difficult to request their participation. We thus asked generally healthy adults who been hospitalized in the past or would be in the future to play the role of the patient. We used the computer-aided facial expression emotion analysis to identify suitable facial expressions, as introduced in Section III.

A. Experimental Conditions

1) Doctor participants: Seven medical students (4 women, 3 men; average age 22.5 years).

2) Patient conditions: Although the actual condition of hospitalized patients varies widely among patients, we had a role-playing patient portray only three conditions, as shown in Figure 3: a patient who feels physically healthy (a "bright patient"), one whose physical condition is unknown (an "expressionless patient"), and one who feels badly and is suffering pain (a "patient in pain").

3) Video recording: The three photographs in Figure 3 were shown to the doctor participants along with the following explanation. "The photographs you will see show a patient you visit during regular morning rounds in a hospital. They are of the same patient, but his condition is different in each picture. In the first one, he feels physically healthy; in the second one, his physical condition is unknown; and in the third one, he feels ill and is suffering pain. After looking at each photograph, "Please greet the patient as a doctor for about 5 seconds or so." This process was repeated 5 days later using three of the students and the same photographs. On the first day, all students greeted the patient as they thought best. Before the process on the second day, the three selected students received coaching based on the authors' previous findings. On the second day, the three students greeted the patient again. We recorded their greetings (30 recordings in total) and used them for our evaluation of facial expressions.

The purpose of this experiment was not to identify the most appropriate facial expressions but simply ones that would be acceptable for most patients, so having a large group of doctors was not needed. In this paper, we consider "appropriate" to mean "acceptable."

Figure 3. Conditions portrayed by role-playing patient.

B. Subjective Evaluation

To make the subjective evaluation more effective, we first had 16 people view and evaluate each video recording and removed the ones, in which the student's facial expression was judged to be unacceptable. We then had 31 other people view and evaluate the remaining recordings.

1) First subjective evaluation: We showed the video recordings to 5 men and 11 women (average age 46) without sound. We asked them to judge whether the doctor's facial expression was appropriate for the situation on a 5-point scale ("1 completely appropriate," "2 somewhat appropriate," "3 neutral," "4 not so appropriate," "5 inappropriate"). We also asked them to comment on anything they felt or noticed. We showed the recordings without sound because we wanted them to focus on appropriate facial expressions in medical communication situations, and emotion is easier to read from speech than from facial expressions. The results are shown in Table I.

2) Second subjective evaluation: For the second subjective evaluation, we eliminated the recordings with a score of 4 or 5 in the first evaluation, except for one score-4 recording, because the facial expressions for those recordings would be unacceptable for patients. One score-4 recording was kept because otherwise there would have been only three recordings for the "patient in pain." The videos used were B-1 to B-6 for the "bright patient" (shown in blue in Table I), E-1, E-2, E-3, E-4, and E-8 for the "expressionless patient" (shown in gray), and P-1 to P-4 and P-8 for the "patient in pain" (shown in yellow).

We showed these videos to 17 men and 14 women (average age 35.9) without sound. We asked them to judge whether the doctor's facial expression was appropriate for the patient's condition on the same 5-point scale. We again asked them to also comment on anything they felt or noticed. The results are shown in Table II. As in the first evaluation, we showed the videos without sound so that the participants would focus their attention on the facial expressions.

TABLE I. RESULTS OF FIRST SUBJECTIVE EVALUATION

	Video-ID	F51	F67	F65	M54	F59	F58	F57	M48	F55	F50	F20	F27	F24	M34	M34	M32
Bright	B-1	1			2			1	2	1	1	2	2	1	1	2	1
	B-2	1	2	2	2	1	1	1	2	1	1	2	1	1	2	2	1
	B-3	1			2			2	2	1	1	2	2	1	2	1	1
	B-4	1	2	1	2	1	1	2	1	2	2	1	2	2	2	2	1
	B-5							1	1	2	1	2	2	1	2	2	2
	B-6							2	2	2	1	2	2	2	2	2	1
	B-7	1			4			3	1	1	1	2	3	2	2	2	1
	B-8	1	2	1	2	1	2	3	2	2	3	2	2	2	4	4	2
	B-9	1	1	2	4	3	2	2	4	2	4	2	2	3	3	2	3
	B-10	2			2			2	4	1	2	2	3	3	4	3	4
Expressionless	E-1	1			2			3	1	2	1	3	3	1	2	1	2
	E-2	1			2			2	2	3	1	2	3	2	2	2	2
	E-3							2	3	2	2	2	2	2	2	2	1
	E-4							2	3	2	2	2	2	2	2	2	1
	E-5	1			2			3	2	1	1	3	4	2	2	2	2
	E-6	1	2	2	2	4	2	2	2	4	2	2	2	2	2	2	1
	E-7	1	2	3	2	1	2	1	4	3	4	3	1	3	3	2	2
	E-8	2	2	1	2	3	1	2	2	2	3	3	2	3	3	3	3
	E-9	2	3	2	3	4	3	2	1	3	2	3	2	3	3	4	1
	E-10	2			2			3	4	1	2	3	3	3	4	2	4
in Pain	P-1							2	2	1	2	2	2	3	2	2	1
	P-2	1			2			2	3	1	1	3	1	3	2	3	2
	P-3							1	1	1	2	2	2	3	3	3	2
	P-4	2			2			1	1	2	2	2	4	2	2	2	3
	P-5	2	2	2	2	2	2	2	3	3	2	1	3	4	2	2	4
	P-6	2	2	2	2	2	2	2	3	3	2	1	3	4	2	2	4
	P-7	2	2	2	4	2	3	2	2	3	2	1	2	4	4	3	3
	P-8	2			2			3	2	3	2	3	3	3	4	2	2
	P-9	1	2	3	4	2	3	4	3	3	3	4	2	5	4	4	1
	P-10	2			4			4	4	2	2	4	3	3	2	3	4

TABLE II. RESULTS OF SECOND SUBJECTIVE EVALUATION

	Video-ID	Score 1	2	3	4	5	avg.
Bright	B-2	9	17	4	1	0	1.9
	B-4	13	8	9	1	0	1.9
	B-3	7	10	12	2	0	2.3
	B-5	2	15	13	1	0	2.4
	B-6	5	8	16	2	0	2.5
	B-1	3	9	16	3	0	2.6
Expressionless	E-1	4	10	14	3	0	2.5
	E-3	5	5	19	2	0	2.6
	E-4	4	6	18	3	0	2.6
	E-8	3	5	22	0	1	2.7
	E-2	1	6	13	11	0	3.1
in Pain	P-4	10	8	9	4	0	2.2
	P-3	5	8	16	2	0	2.5
	P-1	3	9	13	5	1	2.7
	P-8	2	10	8	10	1	2.9
	P-2	0	4	14	11	2	3.4

V. QUANTITATIVE ANALYSIS

In addition to the two subjective evaluations, the emotions represented by the facial expressions in the 16 highest ranked videos were identified using our facial expression emotion analysis system.

A. *Facial expressions appropriate for "bright patient"*

The three top-ranked videos for "bright patient" as subjectively evaluated were analyzed from the point of view of what emotions appear in the facial expressions, as identified by the facial expression emotion analysis. Tables III, IV, and V show the results. Table VI shows the results for video B-10, which was evaluated as low.The cells in the tables corresponding to 0 or more and less than 0.2 are shown in blue, 0.2 or more and less than 0.4 in green, 0.4 or more and less than 0.6 in yellow, 0.6 or more and less than 0.8 in orange, and 0.8 or more in red. A representative facial expression is shown in Figure 4. Images of the students corresponding to videos B-2 and B-4 are not shown because permission could not be obtained.

Figure 4. Facial expression of student corresponding to video B-3.

TABLE III. COMPUTER ANALYSIS RESULTS FOR VIDEO B-2 (TOP-RANKED FACIAL EXPRESSION FOR "BRIGHT PATIENT")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.04
00:00.7	0.95	0.00	0.00	0.00	0.00	0.00	0.00	0.05
00:01.1	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.10
00:01.6	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:02.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:02.8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.9	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:04.4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE IV. COMPUTER ANALYSIS RESULTS FOR VIDEO B-4 (2ND-RANKED FACIAL EXPRESSION FOR "BRIGHT PATIENT")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.83	0.00	0.01	0.01	0.00	0.00	0.00	0.16
00:00.7	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.01
00:01.1	0.98	0.00	0.00	0.01	0.00	0.00	0.00	0.01
00:01.7	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:02.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:02.8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.2	0.97	0.00	0.00	0.02	0.00	0.00	0.01	0.00
00:03.7	0.95	0.00	0.00	0.01	0.00	0.00	0.00	0.04
00:04.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:04.8	0.99	0.00	0.00	0.01	0.00	0.00	0.00	0.00
00:05.4	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:06.0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:06.5	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:06.9	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE V. COMPUTER ANALYSIS RESULTS FOR VIDEO B-3 (3RD-RANKED FACIAL EXPRESSION FOR "BRIGHT PATIENT")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:00.7	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:01.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:02.2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:01.7	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:02.8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.3	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.9	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE VI. COMPUTER ANALYSIS RESULTS FOR VIDEO B-10 (LOW-RANKED FACIAL EXPRESSION FOR "BRIGHT PATIENT")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
00:00.6	0.00	0.00	0.00	0.00	0.03	0.02	0.50	0.45
00:01.1	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.97
00:01.6	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.99
00:02.1	0.00	0.00	0.00	0.00	0.01	0.10	0.03	0.85
00:02.6	0.00	0.00	0.00	0.00	0.06	0.00	0.49	0.44
00:03.1	0.02	0.00	0.02	0.00	0.00	0.03	0.01	0.92
00:03.6	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.98
00:04.1	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.96
00:04.7	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.97

The facial expression for "bright patient" was "constant happiness" (expressed more as a laugh rather than simply a smile) for the three top-ranked videos. The lower evaluated video, B-10, was mostly "neutral," and it was judged as showing "nervousness," "no expression," "scary eye," etc., which explains why it received a low evaluation.

B. *Facial expressions appropriate for "expressionless patient"*

The three top-ranked videos for the "expressionless patient" as subjectively evaluated were analyzed from the point of view of what emotions appear in the facial expressions as identified using our facial expression emotion analysis system. Tables VII, VIII, and IX show the results. Table X shows the results for video E-8, which was evaluated as low.

As shown in Figure 5, the image representing the first half of video E-1 is largely "happiness" (expressed as a smile) and "neutral" in the second half (mainly expressionless). The 2nd- and 3rd-ranked videos were mostly "neutral." The lower evaluated video, E-8, was mostly "happiness," which was judged not to be serious enough. It is thought that this is because medical students cannot judge whether an "expressionless patient" is in a good or bad physical condition due to the lack of expression. Since the patient's condition could be bad, an expression showing "happiness" was judged by some as inappropriate. The "neutral" expressions shown in the 2nd- and 3rd-ranked videos were apparently judged as having little effect on the patient. The top-ranked video showed a natural greeting starting with a smile and then transitioning to "neutral" as the "doctor" learned about the patient's condition, which is considered to be a reasonable explanation for the high evaluation.

C. *Facial expressions appropriate for "patient in pain"*

The three top-ranked videos for "patient in pain" as subjectively evaluated were analyzed from the point of view of what emotions appear in the facial expressions as identified using our facial expression emotion analysis system. Tables XI, XII, and XIII show the results. The expression in the top-ranked video is shown in Figure 6. Table XIV shows the results for video P-10, which was evaluated low.

Figure 5. Facial expression of student corresponding to video E-1 (left: first half; right: second half).

TABLE VII. COMPUTER ANALYSIS RESULTS FOR VIDEO E-1 (TOP-RANKED FACIAL EXPRESSION FOR "EXPRESSIONLESS PATIENT")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.66	0.00	0.00	0.00	0.00	0.00	0.00	0.34
00:00.7	0.71	0.00	0.00	0.00	0.00	0.00	0.00	0.29
00:01.2	0.79	0.00	0.01	0.00	0.00	0.00	0.00	0.20
00:01.6	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.06
00:02.2	0.35	0.00	0.01	0.00	0.00	0.00	0.00	0.63
00:02.7	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.07
00:03.2	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.76
00:03.8	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.70
00:04.0	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.59

TABLE VIII. COMPUTER ANALYSIS RESULTS FOR VIDEO E-3 (2ND-RANKED FACIAL EXPRESSION FOR "EXPRESSIONLESS PATIENT")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.7	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.93
00:01.1	0.00	0.00	0.00	0.00	0.00	0.06	0.02	0.92
00:01.6	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.96
00:02.2	0.00	0.00	0.00	0.00	0.00	0.08	0.01	0.92
00:02.6	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.66
00:00.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
00:03.2	0.00	0.01	0.02	0.00	0.00	0.13	0.00	0.84
00:03.6	0.00	0.00	0.01	0.00	0.00	0.05	0.00	0.94
00:03.0	0.00	0.00	0.01	0.00	0.00	0.06	0.00	0.92

TABLE IX. COMPUTER ANALYSIS RESULTS FOR VIDEO E-4 (3RD-RANKED FACIAL EXPRESSION FOR "EXPRESSIONLESS PATIENT")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.96
00:01.6	0.00	0.00	0.00	0.00	0.02	0.01	0.34	0.63
00:02.1	0.00	0.00	0.00	0.00	0.03	0.01	0.35	0.60
00:02.6	0.00	0.00	0.00	0.00	0.08	0.02	0.37	0.52
00:03.1	0.00	0.00	0.00	0.00	0.05	0.05	0.24	0.65
00:03.6	0.03	0.00	0.02	0.00	0.00	0.03	0.01	0.91
00:04.1	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.96

TABLE X. COMPUTER ANALYSIS RESULTS FOR VIDEO E-8 (LOW-RANKED FACIAL EXPRESSION FOR "EXPRESSIONLESS PATIENT")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.0	0.81	0.00	0.00	0.00	0.00	0.00	0.00	0.19
00:00.5	0.60	0.00	0.01	0.00	0.00	0.00	0.05	0.34
00:01.0	0.81	0.00	0.00	0.01	0.00	0.00	0.09	0.09
00:01.6	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.03
00:02.0	0.95	0.00	0.00	0.00	0.00	0.00	0.01	0.04
00:02.5	0.96	0.00	0.00	0.01	0.00	0.00	0.00	0.03
00:03.0	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.01
00:03.7	0.91	0.00	0.02	0.01	0.00	0.00	0.00	0.07
00:04.0	0.87	0.00	0.04	0.01	0.00	0.00	0.00	0.08

The expressions in the higher ranked videos are mainly "neutral," with "surprise," "sadness," or "fear" gradually appearing in some. There was virtually no expression of "happiness," "anger," or "disgust" except for the third frame of video P-1. In contrast, the expression in the lower ranked video, P-10, was mostly "happiness," resulting in comments such as "the grinning made me feel uncomfortable" and "the doctor seemed to be smiling faintly." These comments explain the low evaluation.

Figure 6. Facial expression of student corresponding to video P-4.

TABLE XI. COMPUTER ANALYSIS RESULTS FOR VIDEO P-4 (TOP-RANKED FACIAL EXPRESSION FOR "PATIENT IN PAIN")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.7	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.93
00:01.1	0.00	0.00	0.00	0.00	0.00	0.06	0.02	0.92
00:01.6	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.96
00:02.2	0.00	0.00	0.00	0.00	0.00	0.08	0.01	0.92
00:02.6	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.66
00:00.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
00:03.2	0.00	0.01	0.02	0.00	0.00	0.13	0.00	0.84
00:03.6	0.00	0.00	0.01	0.00	0.00	0.05	0.00	0.94
00:03.0	0.00	0.00	0.01	0.00	0.00	0.06	0.00	0.92

TABLE XII. COMPUTER ANALYSIS RESULTS FOR VIDEO P-3 (2ND-RANKED FACIAL EXPRESSION FOR "PATIENT IN PAIN")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.96
00:01.6	0.00	0.00	0.00	0.00	0.02	0.01	0.34	0.63
00:02.1	0.00	0.00	0.00	0.00	0.03	0.01	0.35	0.60
00:02.6	0.00	0.00	0.00	0.00	0.08	0.02	0.37	0.52
00:03.1	0.00	0.00	0.00	0.00	0.05	0.05	0.24	0.65
00:03.6	0.03	0.00	0.02	0.00	0.00	0.03	0.01	0.91
00:04.1	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.96

TABLE XIII. COMPUTER ANALYSIS RESULTS FOR P-1 (3RD-RANKED FACIAL EXPRESSION FOR "PATIENT IN PAIN")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.0	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.93
00:00.6	0.01	0.00	0.00	0.00	0.00	0.00	0.06	0.93
00:01.1	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.73
00:01.6	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.98
00:02.1	0.00	0.00	0.00	0.00	0.01	0.09	0.09	0.81
00:02.6	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.94
00:03.2	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.96
00:03.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:04.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
00:04.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
00:05.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
00:05.0	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.98

TABLE XIV. COMPUTER ANALYSIS RESULTS FOR VIDEO P-10 (LOW-RANKED FACIAL EXPRESSION FOR "PATIENT IN PAIN")

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:00.5	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:01.0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:01.5	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.17
00:02.0	0.95	0.00	0.00	0.00	0.00	0.00	0.00	0.05
00:02.5	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.01
00:03.0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00:03.5	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.01
00:04.1	0.58	0.00	0.00	0.00	0.00	0.00	0.00	0.41
00:04.7	0.24	0.00	0.04	0.00	0.00	0.00	0.00	0.72

D. Facial expressions appropriate for "auscultation"

To identify facial expressions appropriate for a doctor performing auscultation, videos were created and evaluated using the following procedure.

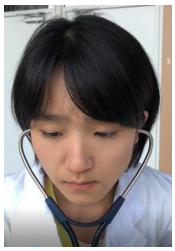
Step 1: Take photograph of patient being auscultated (Figure 7).

Step 2: Record videos of two medical doctors producing facial expressions they thought appropriate for the patient.

Step 3: Analyze videos using computer-aided facial expression emotion analysis system.

Step 4: Have ten potential patients evaluate doctors' facial expressions on 3-point scale (1: appropriate, 2: neutral, 3: inappropriate).

Figure 7. Role-playing patient being auscultated.


(1) Video A-1


(2) Video A-2

Figure 8. Medical students performing auscultation.

TABLE XV. COMPUTER ANALYSIS RESULTS FOR VIDEO A-1.

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:00.6	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.96
00:01.1	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.97
00:01.6	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.93
00:02.1	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.98
00:02.6	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.98
00:03.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:03.6	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.99
00:04.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:04.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:05.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:05.6	0.12	0.01	0.01	0.00	0.00	0.04	0.00	0.81
00:05.6	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.98
00:06.2	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.97
00:06.7	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:07.3	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.98
00:07.8	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.93

TABLE XVI. COMPUTER ANALYSIS RESULTS FOR VIDEO A-2.

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.00	0.03	0.00	0.00	0.00	0.04	0.00	0.92
00:00.6	0.00	0.03	0.01	0.00	0.00	0.01	0.00	0.94
00:01.1	0.00	0.03	0.02	0.00	0.00	0.01	0.00	0.94
00:01.6	0.00	0.03	0.01	0.00	0.00	0.01	0.00	0.94
00:02.1	0.00	0.03	0.01	0.00	0.00	0.02	0.00	0.94
00:02.6	0.00	0.06	0.01	0.00	0.00	0.02	0.00	0.91
00:03.2	0.00	0.07	0.02	0.00	0.00	0.01	0.00	0.90
00:03.7	0.00	0.03	0.03	0.00	0.00	0.04	0.00	0.90
00:04.2	0.00	0.01	0.05	0.00	0.00	0.05	0.00	0.89
00:04.7	0.00	0.01	0.04	0.00	0.00	0.03	0.00	0.91
00:05.2	0.00	0.02	0.02	0.00	0.00	0.03	0.00	0.91
00:05.7	0.00	0.02	0.03	0.00	0.00	0.02	0.00	0.92
00:06.2	0.00	0.02	0.02	0.00	0.00	0.03	0.00	0.93
00:06.7	0.00	0.01	0.03	0.00	0.00	0.02	0.00	0.93
00:07.3	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.96
00:07.7	0.00	0.00	0.01	0.00	0.00	0.03	0.00	0.96
00:08.3	0.00	0.00	0.01	0.00	0.00	0.05	0.00	0.94

TABLE XVII. RESULTS OF SUBJECTIVE EVALUATION.

Video-ID	Score			avg.
	1	**2**	**3**	
A-1	10	0	0	1.0
A-2	9	1	0	1.1

As shown in Tables XV and XVI, the results of the facial expression emotion analysis were mostly “neutral.” As shown in Table XVII, the potential patients felt that the facial expressions in Figure 8 were appropriate.

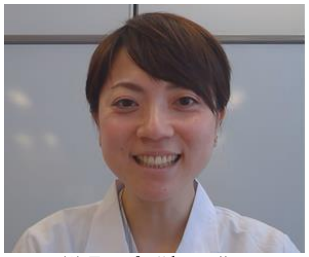
The potential patients felt that the facial expressions in the two videos were mainly "neutral," meaning that they felt that the doctor was performing the auscultation in a serious manner.

VI. GENDER AND AGE-SPECIFIC DIFFERENCES AMONG EVALUATORS

To investigate the difference in evaluation by gender and age of the evaluator, we made videos of a man and a woman playing the role of a doctor and producing expressions in accordance with the findings reported above.

A. Video creation of model doctors

As described in the previous section, acceptable expressions were identified for four situations. For patients who feel physically healthy, the most acceptable facial expression is "continuous happiness" (expressed more as a laugh rather than simply a smile). For patients without a facial expression, the most acceptable facial expression is initially "happiness" (expressed as a smile) and then "neutral" (without expression). For patients feeling ill and suffering pain, the most acceptable facial expression is "neutral" with a little "sadness" or "surprise." For patients being auscultated, the most acceptable facial expression is "continuous neutral" (expressed as serious). We recorded video of the two “doctors” as they produced these expressions (Figures 9–12).


(1) Female “doctor”

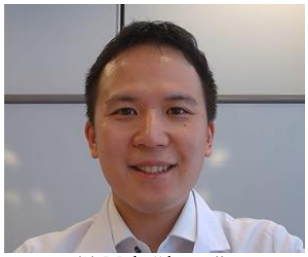
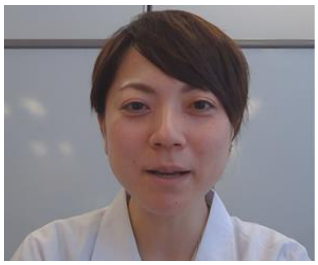
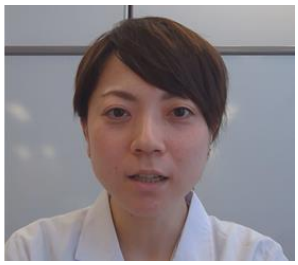
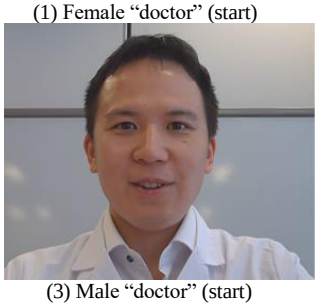

(2) Male “doctor”

Figure 9. Appropriate facial expressions for bright patiets.

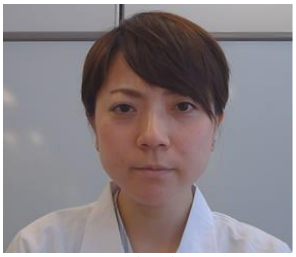

(1) Female “doctor” (start)


(2) Female “doctor” (half-way)


(3) Male “doctor” (start)


(4) Male “doctor” (half-way)

Figure 10. Appropriate facial expressions for expressionless patiets.


(1) Female “doctor”

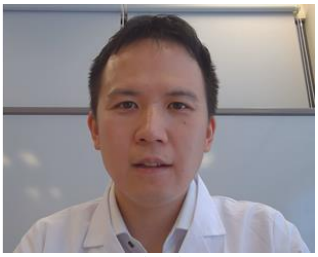

(2) Male “doctor”

Figure 11. Appropriate facial expressions for patiets in pain.


(1) Female “doctor”

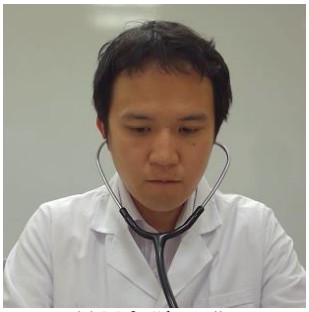

(2) Male “doctor”

Figure 12. Appropriate facial expressions for auscultation.

We showed the videos without sound to 32 men and 47 women (average age 46.7), as detailed in Table XVIII. We asked these evaluators to judge whether the doctor’s facial expression was appropriate for the situation on a 3-point scale ("1: appropriate," "2: neutral," "3: inappropriate").

TABLE XVIII. BREAKDOWN OF EVALUATORS

	20s	30s	40s	50s	60s	70s	Total
Female	7	5	6	11	10	8	47
Male	11	7	5	5	2	2	32
Total	18	12	11	16	12	10	79

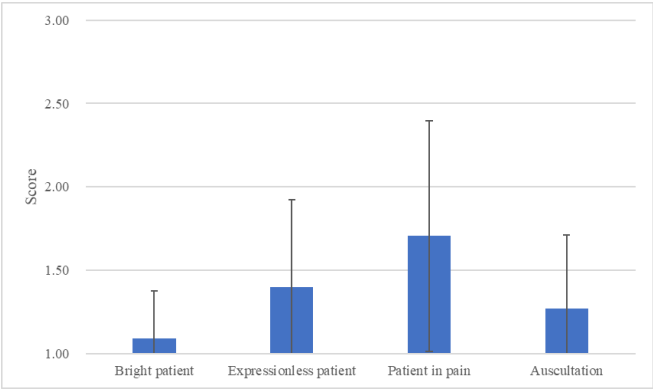


Figure 13. Average score and standard deviation for each condition.

We also asked them to comment on anything they felt or noticed. We showed the videos without sound so that they would focus their attention on the facial expressions. As shown in Figure 13, the expressions for the "bright patient" were judged to be the most appropriate while those for the "patient in pain" were judged to be the most inappropriate. This latter result may be because the “patients” produced a variety of facial expressions for the "patient in pain" situation. This is consistent with the results of the first experiment, in which facial expressions appropriate for doctors were identified (Table I).

B. Differences in evaluation by gender

Looking at the evaluation results in Figure 14 for the "bright patient" by evaluator gender, we see that there was little difference between the male and female evaluators. None of the evaluators gave a score of 3 ("inappropriate") while 91.5% of the female evaluators and 90.3% of the male evaluators gave a score of 1 ("appropriate").

The expressions for the "expressionless patient" were judged quite differently: 71.0% of the male evaluators and 56.4% of the female evaluators gave a score of 1 ("appropriate"), 42.6% of the female evaluators and 25.8% of the male evaluators gave a score of 2 (“neutral"), and 1.1% of the female evaluators and 3.2% of the male evaluators gave a score of 3 (“inappropriate"). In short, the female evaluators were more critical in their evaluations, which conforms to the results of previous studies showing that women are better at reading facial expressions [26]–[30].

The expressions for the "patient in pain" were also judged quite differently, with the female evaluators again being more critical: 56.5% of the male evaluators and 34.0% of the female evaluators gave a score of 1 ("appropriate"), 51.1% of the female evaluators and 32.3% of the male evaluators gave a score of 2 ("neutral)," and 14.9% of the female evaluators and 11.3% of the male evaluators gave a score of 3 ("inappropriate"). These results also conform to the results of previous studies showing that women are better at reading facial expressions [26]–[30].

The expressions for "Auscultation" were judged a bit differently, with the male evaluators being slightly more critical. None of the evaluators gave a score of 3 ("inappropriate"; 76.0% of the female evaluators and 70.7% of the male evaluators gave a score of 1 ("appropriate"). This difference by gender is within an acceptable error range.

These results totally indicate that women are more likely to be sensitive to facial expressions than men.

C. Differences in evaluation by age

The average evaluation scores by evaluator age are plotted in Figures 15–22. The number of evaluators in each age group is shown in Table XVIII.

For the “bright patient,” while there was some variance by age, the male and female evaluators of all ages mostly gave scores of 1 ("appropriate"). Therefore, evaluator age had little effect on the score for the “bright patient.”

For the “expressionless patient,” the female evaluators were somewhat critical overall. Although there was some

variation by age, it is hard to assign a reason for this due to the small number of evaluators in each age group.

For the “patient in pain,” there was a tendency for the individual differences for both the male and female evaluators to be critical. The evaluators in their 20s tended to give better scores (average score of 1.33 vs. overall average score of 1.75). This may be because patients in their 20s are less concerned with a doctor’s facial expression, as reflected in the comment from one of the evaluators in the 20s group: "I have never minded the doctor's facial expression."

For “Auscultation,” the male and female evaluators in their 30s were more critical (average score of 1.64 vs. overall average score of 1.28). They were also more critical for the “expressionless patient” (average score of 1.71 vs. overall average score of 1.41).

These results indicate that age was not an important factor in the evaluations; however, we plan to investigate this further in experiments using more evaluators.

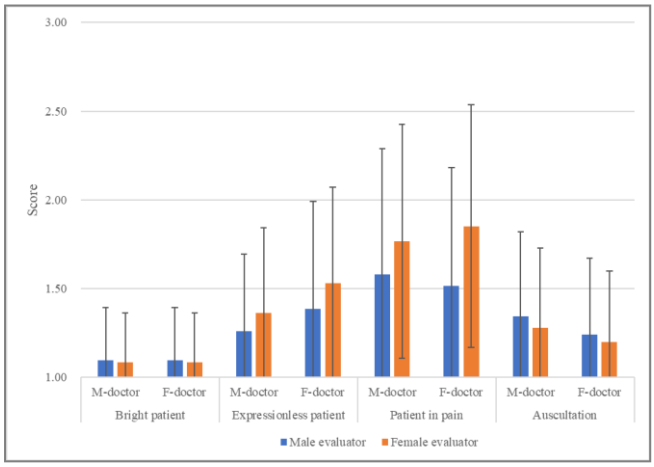


Figure 14. Average evaluation scores and standard deviation by gender.

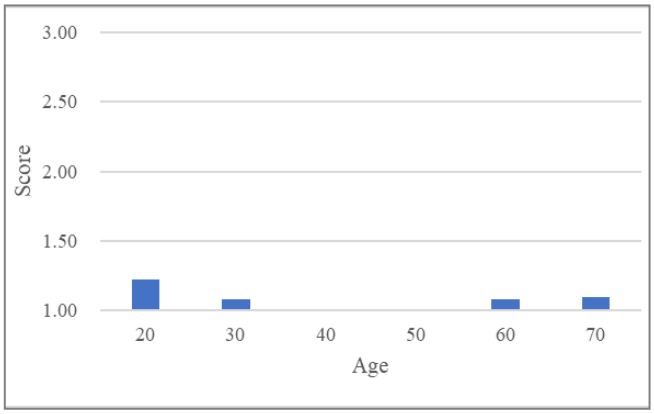


Figure 15. Average scores by age for male evaluators for “bright patient.”

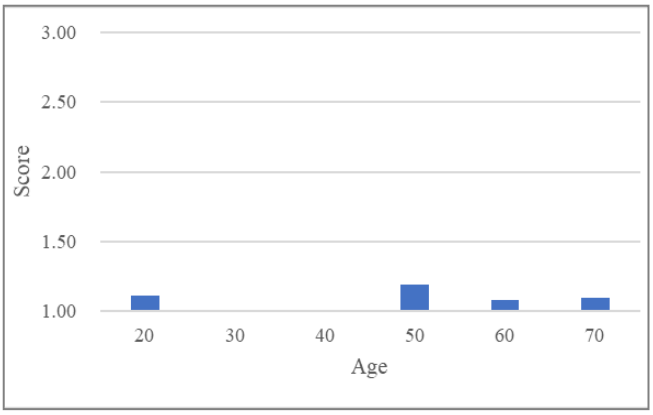


Figure 16. Average scores by age for female evaluators for “bright patient.”

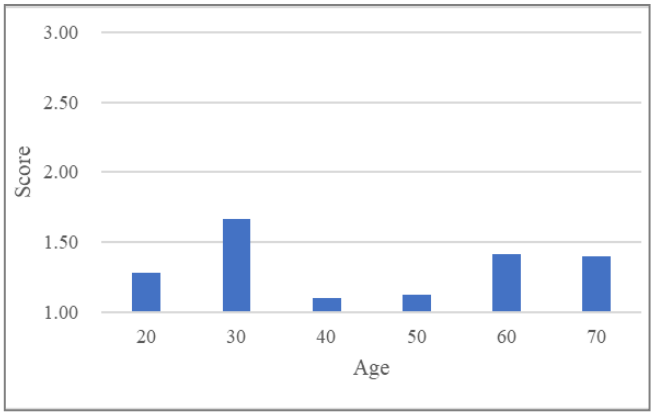


Figure 17. Average scores by age for male evaluators for “expressionless patient.

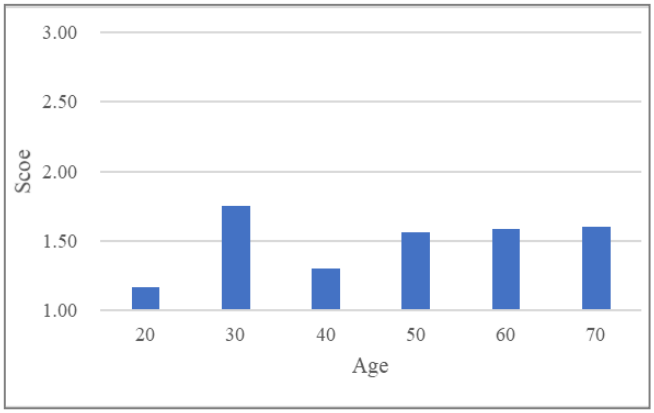


Figure 18. Average scores by age for female evaluators for “expressionless patient.”

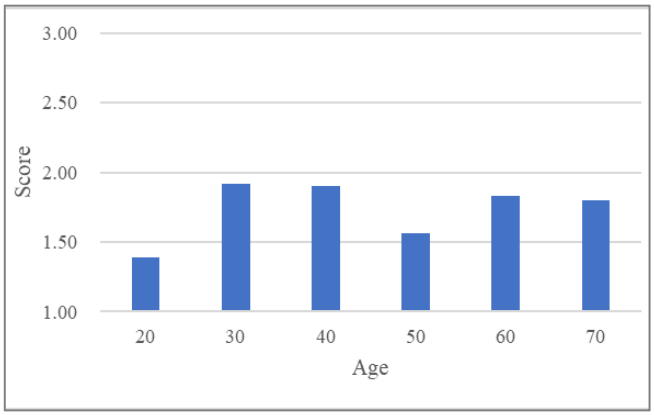


Figure 19. Average scores by age for male evaluators for “patient in pain.”

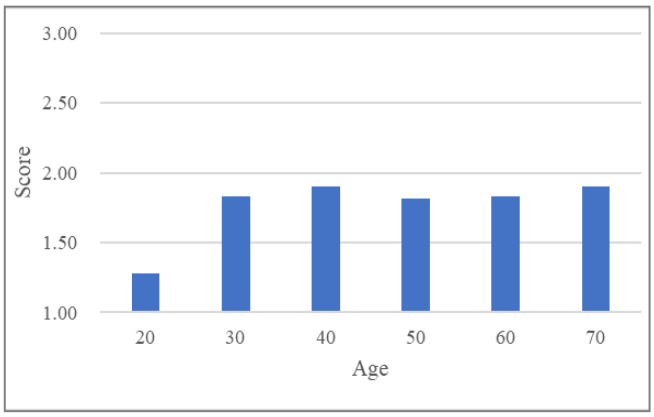


Figure 20. Average scores by age for female evaluators for “patient in pain.”

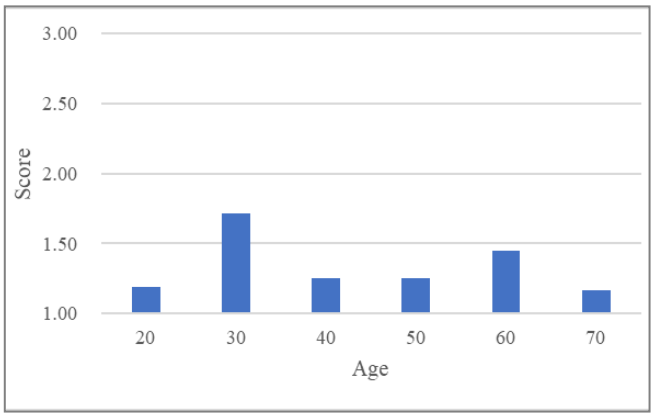


Figure 21. Average scores by age for male evaluators for “Ausculation.”

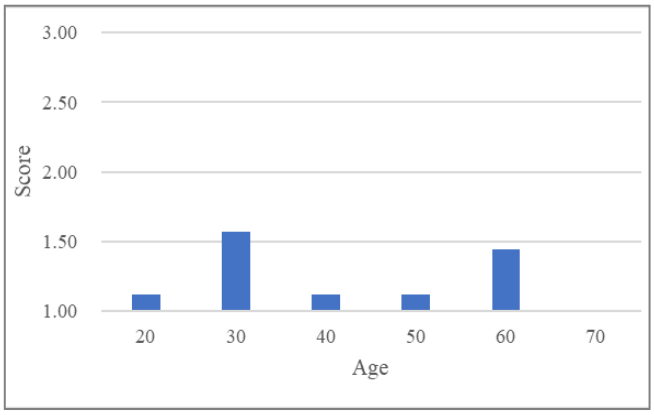


Figure 22. Average scores by age for female evaluators for “Ausculation.”

VII. PEDIATRICIAN'S FACIAL EXPRESSIONS

The expressions appropriate for pediatric patients likely differ from ones appropriate for adult patientss. Due to an influenza outbreak, we were able to get cooperation from only two doctors in this part of our study. We used our computer-aided facial expression emotion analysis system to analyze the facial expressions they produced when shown photographs of pediatric patients in three different conditions.

A. *Experimental procedure*

To identify facial expressions appropriate for greeting hospitalized pediatric patients in the same three conditions described above ("bright patient," "expressionless patient," and "patient in pain"), we used the following procedure.

Step 1: We photographed role-playing pediatric patients as they produced expressions for each of the conditions. (The photographs of the 8-year-olds are not included here for privacy reasons.)

Step 2: We recorded video of the two doctors as they produced facial expressions they thought appropriate for each photograph.

Step 3: We analyzed the emotion shown by their expressions by using computer-aided facial expression emotion analysis.

B. *Results*

The results of the facial expression emotion analysis for the two pediatricians are shown in Tables XIX–XXI.

For the "bright patient," the appropriate facial expression for adult patients was "happiness" while the pediatricians produced expressions covering both "happiness" and "neutral," which created a pleasant smiley feeling.

For the "expressionless patient," the appropriate expressions for adult patients were first "happiness" and then "neutral," whereas the pediatricians produced "neutral" expressions.

For the "patient in pain," the appropriate facial expression for adult patients was "neutral," whereas the pediatricians first produced a "neutral" expression and then a "happiness" expression. Thus, for pediatric patients with a

bright expression, the pediatricians greet them with a natural expression. For expressionless pediatric patients, they are expressionless, like the patient. Unlike doctors for adults, the most important facial expression for pediatric physicians was found to be the final encouraging smile as part of speaking naturally to pediatric patients in pain.

TABLE XIX. COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "BRIGHT PATIENT"

(1) PEDIATRICIAN_1

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.0	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.96
00:00.5	0.12	0.00	0.02	0.01	0.00	0.02	0.01	0.83
00:01.0	0.12	0.00	0.03	0.00	0.00	0.00	0.00	0.84
00:01.5	0.54	0.00	0.01	0.00	0.00	0.00	0.01	0.43
00:02.0	0.58	0.00	0.02	0.01	0.00	0.01	0.01	0.36
00:02.5	0.06	0.00	0.01	0.00	0.00	0.01	0.02	0.90
00:03.0	0.08	0.00	0.01	0.00	0.00	0.01	0.02	0.87
00:03.5	0.15	0.00	0.01	0.00	0.00	0.01	0.01	0.82
00:04.0	0.49	0.00	0.01	0.01	0.00	0.01	0.01	0.48
00:04.5	0.60	0.00	0.01	0.01	0.00	0.00	0.01	0.37
00:05.0	0.15	0.00	0.07	0.01	0.00	0.01	0.02	0.75
00:05.5	0.24	0.00	0.01	0.00	0.00	0.00	0.01	0.74
00:06.0	0.37	0.00	0.01	0.00	0.00	0.01	0.00	0.60

(2) PEDIATRICIAN_2

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.5	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.98
00:00.9	0.06	0.00	0.01	0.00	0.00	0.01	0.31	0.61
00:01.5	0.03	0.01	0.02	0.01	0.00	0.02	0.05	0.85
00:02.1	0.22	0.00	0.05	0.01	0.00	0.02	0.01	0.69
00:02.6	0.30	0.00	0.02	0.00	0.00	0.01	0.00	0.67
00:00.0	0.00	0.00	0.01	0.02	0.00	0.19	0.01	0.78
00:03.1	0.30	0.00	0.01	0.00	0.00	0.00	0.01	0.68
00:03.6	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.50

TABLE XX. COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "EXPRESSIONLESS PATIENT"

(1) PEDIATRICIAN_1

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.96
00:00.6	0.08	0.00	0.01	0.00	0.00	0.00	0.00	0.91
00:01.1	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.92
00:01.6	0.03	0.00	0.01	0.00	0.00	0.01	0.00	0.95
00:02.1	0.04	0.00	0.00	0.00	0.00	0.02	0.01	0.92
00:02.6	0.02	0.00	0.01	0.00	0.00	0.01	0.00	0.96
00:03.1	0.11	0.00	0.01	0.00	0.00	0.01	0.01	0.87
00:03.6	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.93
00:04.1	0.10	0.00	0.01	0.00	0.00	0.01	0.02	0.85
00:04.6	0.10	0.00	0.00	0.00	0.00	0.01	0.02	0.86
00:05.1	0.10	0.00	0.01	0.00	0.00	0.00	0.01	0.87
00:05.6	0.04	0.00	0.01	0.01	0.00	0.04	0.01	0.88
00:06.1	0.06	0.00	0.01	0.00	0.00	0.01	0.01	0.90

(2) PEDIATRICIAN_2

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.05	0.00	0.01	0.00	0.00	0.01	0.00	0.93
00:00.6	0.07	0.00	0.01	0.00	0.00	0.01	0.00	0.92
00:01.1	0.06	0.00	0.01	0.01	0.00	0.01	0.01	0.90
00:01.6	0.01	0.00	0.01	0.00	0.00	0.04	0.01	0.92
00:02.1	0.02	0.00	0.01	0.00	0.00	0.02	0.01	0.93
00:02.7	0.01	0.00	0.01	0.00	0.00	0.04	0.01	0.93
00:03.3	0.02	0.01	0.01	0.02	0.00	0.03	0.01	0.89
00:03.7	0.00	0.01	0.03	0.03	0.00	0.12	0.00	0.81
00:04.2	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.98

TABLE XXI. COMPUTER ANALYSIS RESULTS FOR PEDIATRICIANS' FACIAL EXPRESSIONS FOR "PATIENT IN PAIN"

(1) PEDIATRICIAN_1

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.0	0.08	0.00	0.01	0.00	0.00	0.01	0.01	0.90
00:00.5	0.02	0.00	0.00	0.00	0.00	0.01	0.00	0.97
00:01.0	0.05	0.00	0.01	0.00	0.00	0.01	0.01	0.93
00:01.5	0.05	0.00	0.01	0.01	0.00	0.01	0.01	0.93
00:02.0	0.04	0.00	0.03	0.00	0.00	0.00	0.01	0.92
00:02.5	0.08	0.00	0.01	0.00	0.00	0.00	0.01	0.90
00:03.0	0.26	0.00	0.02	0.01	0.00	0.02	0.01	0.69
00:03.5	0.17	0.00	0.01	0.00	0.00	0.01	0.01	0.80
00:04.0	0.45	0.00	0.02	0.01	0.00	0.00	0.00	0.52
00:04.5	0.50	0.00	0.01	0.01	0.00	0.00	0.00	0.47
00:05.0	0.53	0.00	0.01	0.01	0.00	0.00	0.00	0.44
00:05.5	0.65	0.00	0.01	0.01	0.00	0.00	0.00	0.32
00:06.0	0.43	0.00	0.02	0.01	0.00	0.01	0.01	0.52
00:06.0	0.45	0.00	0.03	0.02	0.00	0.02	0.00	0.48

(2) PEDIATRICIAN_2

Time	Happiness	Anger	Contempt	Disgust	Fear	Sadness	Surprise	Neutral
00:00.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99
00:00.6	0.03	0.01	0.01	0.01	0.00	0.01	0.13	0.80
00:01.1	0.29	0.00	0.04	0.00	0.00	0.00	0.00	0.66
00:01.6	0.01	0.01	0.07	0.03	0.00	0.06	0.03	0.79
00:02.2	0.02	0.00	0.01	0.01	0.00	0.01	0.02	0.94
00:02.7	0.03	0.00	0.01	0.01	0.00	0.06	0.00	0.88
00:03.3	0.25	0.00	0.01	0.00	0.00	0.00	0.01	0.73
00:03.8	0.07	0.00	0.03	0.00	0.00	0.04	0.00	0.86
00:04.3	0.21	0.00	0.01	0.00	0.00	0.00	0.00	0.76
00:04.8	0.10	0.00	0.04	0.01	0.00	0.03	0.00	0.82
00:05.4	0.83	0.00	0.01	0.00	0.00	0.08	0.00	0.09
00:06.0	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.07
00:06.5	0.98	0.00	0.00	0.00	0.00	0.00	0.00	0.02

VIII. CONCLUSION

Our quantitative analysis of medical student facial expressions when greeting an adult patient to be medically evaluated in the general ward of a hospital revealed acceptable facial expressions. For patients who feel physically healthy, the most acceptable facial expression is "continuous happiness" (expressed more as a laugh than simply as a smile). For patients without a facial expression, the most acceptable facial expression is initially "happiness" (expressed as a smile) and then "neutral" (without expression). For patients suffering pain, the most acceptable facial expression is "neutral" with a little "sadness" or "surprise." During auscultation, the most acceptable facial expression is always "neutral."

There was no significant difference in facial expression required by the doctor depending on the gender and age of the patient.

Analysis of the facial expressions produced by two pediatricians showed that they produced expressions covering both "happiness" and "neutral" when greeting pediatric patients who had a bright expression and produced a neutral expression when greeting expressionless pediatric patients. Most importantly, they need to encourage pediatric patients in pain by ending with a smile.

Although appropriate facial expressions may differ between doctors in different countries due to cultural differences, we believe that the proposed method for identifying appropriate facial expressions is useful in any country.

REFERENCES

[1] A. Oe and Y. Murata, "Quantitative Analysis of Facial Expression for Medical Doctors," IARIA, Proceedings of Global Health 2018, pp. 8–13, 2018.

[2] A. Donabedian, The Definition of Quality and Approaches to Its Assessment. Ann Arbor, MI: Health Administration Press, 1980.

[3] S. H. Kaplan, S. Greenfield, and W. E. Ware, Jr, "Assessing the effects of physician-patient interactions on the outcomes of chronic disease," Medical Care, vol. 27, pp. S110–S127, March 1989.

[4] L. M. Noble, A. Kubacki, J. Martin, and M. Lloyd, "The effect of professional skills training on patient-centredness and confidence in communicating with patients," Medical Education, vol. 41, pp. 432–440, May 2007.

[5] J. A. Hall, D. L. Roter, and N. R. Katz, "Meta-analysis of correlates of provider behavior in medical encounters," Medical Care, vol. 26, pp. 657–675, July 1988.

[6] J. A. Hall, J. A. Harrigan, and R. R. Rosenthal, "Nonverbal behavior in clinician-patient interaction," Applied & Preventive Psychology, vol. 4, pp. 21–37, Winter 1995.

[7] L. M. Ong, J. C. de Haes, A. M. Hoos, and F. B. Lammes, "Doctor-patient communication: A review of the literature," Social Science & Medicine, vol. 40, pp. 903–918, April 1995.

[8] T. L. Thompson, "Interpersonal communication and health care," in Handbook of Interpersonal Communication, 2nd ed., M. L. Knapp and G. R. Miller, Eds. Thousand Oaks, CA: Sage, pp. 696–725, 1994.

[9] S. Williams, J. Weinman, and J. Dale, "Doctor-patient communication and patient satisfaction: A review," Family Practice, vol. 15, pp. 480–492, October 1998.

[10] M. R. DiMatteo, A. Taranta, H. S. Friedman, and L. M. Prince, "Predicting patient satisfaction from physicians' nonverbal communication skills," Medical Care, vol. 18, pp. 376–387, April 1980.

[11] M. R. DiMatteo, R. D. Hays, and L. M. Prince, "Relationship of physicians' nonverbal communication skill to patient satisfaction, appointment noncompliance, and physician workload," Health Psychology., vol. 5, pp. 581–594, 1986.

[12] J. Bird and S. A. Cohen-Cole, "The three-function model of the medical interview. An educational device," Advances in Psychosomatic Medicine, vol. 20, pp. 65–88, 1990.

[13] H. de Haes and J. Bensing, "Endpoints in medical communication research, proposing a framework of functions and outcomes," Patient Education and Counseling, vol. 74, pp. 287–294, March 2009.

[14] Z. Ben-Sira, "Affective and instrumental components in the physician-patient relationship: An additional dimension of interaction theory," Health and Social Behavior, vol. 21, pp. 170–180, June 1980.

[15] M. K. Buller and D. B. Buller, "Physicians' communication style and patient satisfaction," Journal of Health and Social Behavior, vol. 28, pp. 375–388, December 1987.

[16] B. M. Korsch and V. F. Negrete, "Doctor-patient communication," In Physician-patient Communication: Readings and Recommendations, G. Henderson, Ed. Springfield, IL: Charles C. Thomas, pp. 29–40, 1981.

[17] S. J. Williams and M. Calnan, "Convergence and divergence: Assessing criteria of consumer satisfaction across general practice, dental and hospital care settings," Social Science & Medicine, vol. 33, pp. 707–716, 1991.

[18] R. C. Wasserman, T. S. Inui, R. D. Barriatua, W. B. Carter, and P. Lippincott, "Pediatric clinicians' support for parents makes a difference: An outcome-based analysis of clinician-parent interaction," Pediatrics, vol. 74, pp. 1047–1053, December 1984.

[19] H. S. Friedman, M. R. DiMatteo, and A. Taranta, "A study of the relationship between individual differences in nonverbal expressiveness and factors of personality and social interaction," Journal of Research in Personality, vol. 14, pp. 351–364, September 1980.

[20] J. Bensing, "Doctor-patient communication and the quality of care," Social Science & Medicine, vol. 32, pp. 1301–1310, 1991.

[21] A. Mehrabian, "Communication without words," Psychol. Today, vol. 2, pp. 52–55, 1968.

[22] R. L. Birdwhistell, Kinesics and Context: Essays on Body Motion Communication. Philadelphia: University of Pennsylvania Press, 1970.

[23] H. S. Friedman, "Nonverbal communication between patients and medical practitioners," J. Soc. Issues, vol. 35, pp. 82–99, Winter 1979.

[24] M. R. DiMatteo, L. M. Prince, and R. Hays, "Nonverbal communication in the medical context: The physician-patient relationship," In Nonverbal Communication in the Clinical Context, P. D. Blanck, R. Buck, and R. Rosenthal, Eds. University Park, PA: Pennsylvania State University Press, pp. 74–98, 1986.

[25] W. B. Gudykunst and Y. Y. Kim, "Communicating with strangers: An approach to intercultural communication (3rd ed.)," New York: McGraw-Hill, 1997.

[26] K. Shimoda, M. Argyle, and P. E. Ricci Bitti, "The intercultural recognition of emotional expressions by three national racial groups: English, Italian and Japanese," European Journal of Social Psychology, vol. 8, pp. 169–179, 1978.

[27] J. A. Hall, "Gender effects in decoding nonverbal cues," Psychological Bulletin, vol. 85, pp. 845–857, 1978.

[28] G. Kirouac and F. Y. Doré, "Accuracy of the judgment of facial expression of emotions as a function of sex and level of education," Journal of Nonverbal Behavior, vol. 9, pp. 3–7, 1985.

[29] S. J. Nowicki and M. Hartigan, "Accuracy of facial affect recognition as a function of locus of control orientation and anticipated interpersonal interaction," The Journal of Social Psychology, vol. 128, pp. 363–372, 1988.

[30] B. Montagne, R. P. C. Kessels, E. Frigerio, E. H. F. de Haan, and D. I. Perrett, "Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity?" Cognitive Processing, vol. 6, pp. 136–141, 2005.

[31] E. Hampson, S. M. van Anders, and L. I. Mullin, "A female advantage in the recognition of emotional facial expressions: test of an evolutionary hypothesis," Evolution and Human Behavior, vol. 27, pp. 401–416, 2006.

[32] D. M. Isaacowitz, C. E. Löckenhoff, R. D. Lane, R. Wright, L. Sechrest, R. Riedel, and P. T. Costa, "Age differences in recognition of emotion in lexical stimuli and facial expressions," Psychology and Aging, vol. 22, pp. 147–159, 2007.

[33] T. Ruffman, J. D. Henry, V. Livingstone, and L. H. Phillips, "A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging," Neuroscience and Biobehavioral Reviews, vol. 32, pp. 863–881, 2008.

[34] L. H. Phillips and R. Allen, "Adult aging and the perceived intensity of emotions in faces and stories," Aging Clinical and Experimental Research, vol. 16, pp. 190–199, 2004.

[35] Cognitive Services, Microsoft Azure, https://azure.microsoft.com/en-us/services/cognitive-services/ [retrieved: June, 2019].

Sensitive Data Discovery in Care Pathways using Business Process Modelling and HL7-CDA

Intidhar Essefi
University of Tunis el Manar, The Higher Institute of Medical Technologies of Tunis
Research Laboratory of Biophysics and Medical Technologies
Tunis, Tunisia
e-mail: essefi.intidhar@gmail.com

Hanen Boussi Rahmouni
University of the West of England, the Computer Science Research Center
Bristol, UK
e-mail: hanene4.rahmouni@uwe.ac.uk

Mohamed Fethi Ladeb
Radiology Department
Kassab Orthopedics Institute
Manouba, Tunisia
e-mail: fethiladeb@hotmail.fr

***Abstract*— Medical data communication is an important process enabling collaboration between healthcare professionals. The use of patient Electronic Health Record (EHR) ensures an enhanced continuity of care since it provides a centralized patient information access in a seamless way. In data protection law, the electronic exchange of medical data should comply with privacy obligations and data security safeguards. It is therefore a legal requirement for hospitals to ensure that patient information are processed and shared throughout clinical business processes in a standardized and structured form in order to be able to clearly discover and highlight the patient protected information manipulated in each process. In this work, we propose a clinical pathway specification methodology that is at the same time data driven and privacy aware. Our model gives special attention to the structure and the data content of shared medical documents. These documents are usually structured following the Health Level 7 (HL7)-Clinical Document Architecture (CDA). This research aims to define a legally shared HL7 structure of medical data ought to be processed and exchanged within clinical processes. Throughout the use of an ontology information model of medical data, we are aiming to capture and classify the data used in each clinical process into categories. We put special emphases on the level of protection required by each category of data in respect of the international health data legislation namely, the Health Insurance and Accountability Act (HIPAA). Particularly, our model ensures the implementation of the privacy by design principle since it ensures the adoption of data protection requirements starting from a very early stage of Hospital Information Systems (HIS) design.**

Keywords-business process modeling; clinical pathways; data driven; legally HL7 structured medical data; HIPAA legislation; ontology; patient privacy; privacy requirements; CDA.

I. INTRODUCTION

In the era of EHRs (Electronic Health Records), many questions are raised: who can access these records in the course of providing health care to the patient, under which circumstances and how? When, if so, can these documents be used for other purposes, such as public health, care quality improvement and public research? When is the consent of the data subject necessary for such "secondary" use of the patient's data? Without the consent of the patient, what are the implications for confidentiality and the sharing of personal medical records or their secondary use?

In order to improve services' quality within a hospital environment, it is important to automate the underlining workflows of each clinical process adopted within the hospital. This requires an explicit design and the implementation of business process models tailored for the concerned field. Medical healthcare is a multidisciplinary field. Its business processes and workflows are very complex. Throughout each medical process, several types of clinical information need to be circulated and treated within or without the hospital's boundaries. Medical data are produced, transmitted between medical departments and shared between healthcare professionals as specified by the clinical pathways enforced by the hospital information systems in use. Throughout this processes, several types of medical data documents are being manipulated, including admission papers, insurance documents, prescriptions, confidential letters, medical images, imaging reports, biological reports, other types of medical reports, etc. All the mentioned clinical documents include diverse health information, among which we distinguish sensitive information that is considered as highly Protected Health Information (PHI) [1]. Personal healthcare information is not only used in healthcare practices and shared between healthcare professionals, but also in public practices and research activities such as public health surveillance and public health research. Public health practices and research present risks that are related to the unauthorized disclosure of PHI [2]. Therefore, it is crucial for healthcare organizations to ensure PHI protection and to preserve the privacy of individuals. Particularly, the individual's privacy protection is required by legislation, such as the Health Insurance Portability and Accountability Act (HIPAA) legislation [3][4] and the GDPR (General Data Protection Regulation) [5] on personal data protection. Consequently, privacy requirements should be respected and ensured when designing systems and procedures for health data management. For that, data oriented care pathway models

present a highly required mean of sensitive data discovery within the shared clinical documents. Furthermore, using a standard structure of medical documents facilitates the discovery and the protection of the included PHI. Medical data classification using ontology for the clinical documents architecture could provide more data fluidity in order to implement personal data protection law. This facilitates the map of each level of protection to a set of privacy requirements as demanded by international health data legislation namely, the Health Insurance and Accountability Act (HIPAA). We believe our model could enhance compliance with privacy with regards to the protection of sensitive patient data within hospitals. Furthermore, the model we are presenting in this paper, describes the different clinical workflows typically included in hospitals care pathways as well as the set of input and out-put data for each workflow, which simplifies the sensitive data discovery task.

Our approach suggests a set of steps towards a legally shared HL7 clinical documents architecture ontology based on the privacy by design principle, which means the implementation of privacy requirements since an early stage of healthcare information systems design with respect to carrying out clinical pathways. In this paper, we take the osteosarcoma clinical pathway as a case study to validate our approach. The details of our approach are as follows:

- Model medical care pathways as business processes that emphasise shared clinical data aiming to identify sensitive health information among them.
- Identify the privacy requirements and procedures for each type of sensitive health data identified within the business process representing each care pathway.
- Identify a clinical data model based on the business process modelling.
- Define the sensitive health information categories.
- Define the HIPAA legislation requirements to preserve the patient's privacy and confidentiality with regards to the use of their PHI.
- Model the clinical pathways based on business process modelling in order to extract the shared clinical documents between healthcare professionals.
- Map clinical documents' data to a set of attributes as a step towards a clinical data ontology development.
- Identify the PHI underlining each process model.
- Define privacy requirements for PHI protection from any disclosure or misuse.

This paper is divided into sections as follows: in Section II, we present the related work; in Section III, we present the clinical pathway subject to study, as well as the clinical pathway modelling language of our choice. We adopt a data driven business process clinical pathway modelling approach. In Section IV, we present our clinical document architecture. In Sections V and VI, we define the privacy requirements for PHI, followed by results and discussion. In Section VII, we present the conclusion and future work.

II. RELATED WORK

In the literature, the study of clinical care pathways is vast due to the diversity of the modelling approaches, the analyses purposes and the care pathway level of specification. Depending on the concerned field, four groups of modelling techniques are defined [6][7][8]: 1) *Statistical and mathematical techniques*, based on the finding of significant relations among variables in order to assess the relation between patient's identifiers and their medical history [8], 2) *Data mining for clinical pathways*, based on discovering patterns in the medical events sequencing in order to predict the outcome of the next step of the care pathway [7], 3) *Business process modelling*, the modelling stage relies on a domain specific ontology for presenting systems in order to be analyzed and improved, it gives attention to the medical resources and documentation presentation in addition to the control-flow [9] [10] and 4) *Process mining algorithms*, dedicated to the processes analyses [11].

The modelling approach is based on questionnaires collecting information throughout interviews with doctors. They do not consider any data source and are subjective. For that, clinical pathways are textually and medically described. Their business processes and workflows are mostly detailed by doctors using textual description of the sequenced tasks. Due to the technological revolution in the medical field that includes medical information systems, several methods and business process modelling languages were suggested. This includes the Integration DEFinition language (IDEF) (V.0 and V.3), the Unified Modelling Language (UML) V.2.0 and the Business Process Model and Notation language (BPMN). Most of these technologies were also used to model clinical pathways [12].

BPMN is the most widely used and accepted language in medical process modelling thanks to its simple and high-level process construction. Clinical pathways processes are known as complex to be understood by medical practitioners. This needs a transparency of the whole process elements such as structures, participants, tasks, roles, etc. Modelling care pathways in the form of clinical processes is considered as a solution to overcome its complexity and define its requirements with regards to patients and health care professionals' needs. Therefore, medical process models should be simple, transparent and understandable as much as possible [13][14].

Despite the importance of business process modelling in clinical pathways and efforts for processes' automation, few works are dealing with care pathways' automation. Most of them are relying on a business process-based modelling approach. Besides, there is some suggested BPMN extension implementation such as the Clinical Pathway (CP) extension of the BPMN, called BPMN4CP, which proposes an ontology based- extension for e-health process

management. Other existing research works offer clinical textual description of the care pathways processes. In addition, other works were interested in analyzing systems behavior throughout business process-based modelling using UML, particularly, UML class diagram [15][16][17].

However, less effort has been made in investigating approaches for clinical data modelling with special interest in privacy preservation. In this context, our work is addressed to the respect of privacy requirements since a very early stage of HIS design in order to ensure a protected rolling of data driven business processes that are clinical pathway-oriented. Based on the modelling approach and the implementation of the *Business Process Modelling* technique, we highlight shared medical data throughout document centric clinical care pathway modelling and HL7-CDA standard investigation [18]. This is for presenting a shared data model and mapping them to a set of characteristics aiming to develop an ontology representing data related to the care pathway field. This will facilitate the definition of the required security level for each data type with respect to personal data protection legislation.

III. CLINICAL PATHWAYS

Clinical pathways are acknowledged as complex processes due to the diversity of the participating entities (e.g., healthcare professionals and medical service providers). Throughout the literature exploration of the clinical pathways' modelling and automation, we extracted the main phases underlining care pathway processes. A generic clinical pathway begins by an admission phase in which the patient is allowed to get access to the care establishments. This is usually followed by a diagnosis phase: that describes the visiting of the consulting doctor and having clinical diagnosis performed. The treatment phase should then occur: after identifying the pathology in the second phase, the treating doctor identifies the treatment protocol. This clinical pathway ends with the follow-up phase which allows the involved practitioner to monitor and evaluate the effectiveness of the prescribed treatment or to control the pathology progression [19].

The clinical pathway is a set of processes and sub-processes in which one or more healthcare professionals participate. The business process modelling of clinical pathways allows to identify the tasks, the participants and their roles in the care pathway proceeding. Even the shared data between healthcare professionals may also be modelled and identified among a clinical pathway-oriented business process [20].

In the following sections, we will detail the clinical care pathway of osteosarcoma and describe a step by step methodology to model our clinical business process and sensitive data discovery.

A. An Overview of Osteosarcoma Clinical Pathways

Osteosarcoma is a bone cancer. It most commonly reaches those aged from 10 to 30. A great part of this affected population concerns teenagers. Each year, from 800 to 900 people are estimated to be diagnosed with osteosarcoma in the United States. Osteosarcomas are primary malignant bone tumors. They can be classified according to cells' behavior under the microscope as high, intermediate or low grade [21].

Osteosarcoma clinical care pathways are characterized by their complex and multidisciplinary procedures with their difficult management facts. By Ferrante [13], the osteosarcoma first diagnosis starts with symptoms appearances like bone pain or soreness, a felt mass through the skin, swelling and redness, etc. During the clinical pathway diagnosis phase, while an osteosarcoma is suspected some standard imaging exams must be performed. Once the osteosarcoma is confirmed and its malignancy is not excluded, a biopsy should be performed allowing the cancer staging. As a final checkup step, several imaging exams are performed to verify the existence of metastases. A percentage of 85% indicates that the most common metastases appear in the lung whereas the bone is considered as the second most common site of distant disease [22].

The osteosarcoma checkup and grading steps allow the choice of the treatment procedure which includes chemotherapy, radiation therapy and surgery operation. The identification of the right therapy protocol is based on biological analyses. To verify and evaluate the treatment efficiency, the patient has to be periodically followed-up. This osteosarcoma clinical pathway follow-up step is based on the performing of imaging exams in addition to biological analyses as needed [22][23][24].

The complexity of the osteosarcoma clinical pathway business process is due to the collaboration between healthcare professionals from several medical departments. This process cannot be accomplished without clinical data sharing and transmission. Among those data, we find various sensitive data which are individually identifiable that are transmitted or maintained in electronic media [25]. For that, it is necessary to respect the applicable data protection regulation. For that, we need data driven clinical pathways' models, which facilitate the data discovery step and mapping them to sensitive data protection principals. Those models allow defining shared and transmitted clinical documents throughout medical processes. The clinical care pathways modelling step is based on the clinical processes description within the literature while the data discovery step is based on the investigation of the HL7-CDA standard. Then, we defined sensitive data according to the HIPAA regulation. As a consequence, the sensitive data discovery step is fulfilled according to both HL7-CDA standard and HIPAA regulation.

B. BPMN As Clinical Pathway Modelling Language

To model clinical pathways, we used the BPMN as a modelling language. It is the most widely used language in healthcare business process modelling. First, we explored

the clinical healthcare pathways in the literature on description of clinical pathways. Then, we divided them into three main phases, diagnosis or check-up, treatment and follow-up, mentioned above. Throughout the studied clinical pathways, we present the osteosarcoma clinical pathway as a case study to illustrate our data driven clinical healthcare pathway model. In order to elaborate the clinical pathway data driven model, we used the common patterns and symbols of the BPMN modelling language [12].

C. Osteosarcoma Clinical Pathway Modelling Using BPMN

In order to identify the clinical data that may be transmitted and shared between healthcare professionals as required by standard care pathway specifications, we modelled the osteosarcoma clinical pathway in the form of a business process model. In this way, we could first identify the characteristics of performed clinical tasks. Then, we could highlight the data driven tasks for the chosen pathology. The sections below present the data driven clinical pathways of osteosarcoma for the check-up, the treatment and the follow-up phases respectively.

1) Osteosarcoma checkup clinical pathway

Fig. 1 presents the check-up phase of the osteosarcoma clinical pathway as well as the shared and transmitted clinical documents, ensuring the steps required by the care pathway definition. The osteosarcoma clinical pathway is complex and involves collaboration of healthcare providers and diverse medical services including radiology, biology, nuclear medicine, surgical units, etc., as shown in Fig. 1, Fig. 2 and Fig. 3. After the patient admission, a medical consultation takes place. According to the clinical examination, medical tests are performed to accomplish the diagnoses phase and precise the pathology Fig. 1. The fulfilment of this phase needs the transmission and the share of diverse medical documents types (e.g., reports, orders, images, etc.) and subtypes (e.g., imaging report, anatomic pathology report, etc.) between healthcare providers. According to the HL7-CDA standard, each clinical document has a predefined standard structure [18].

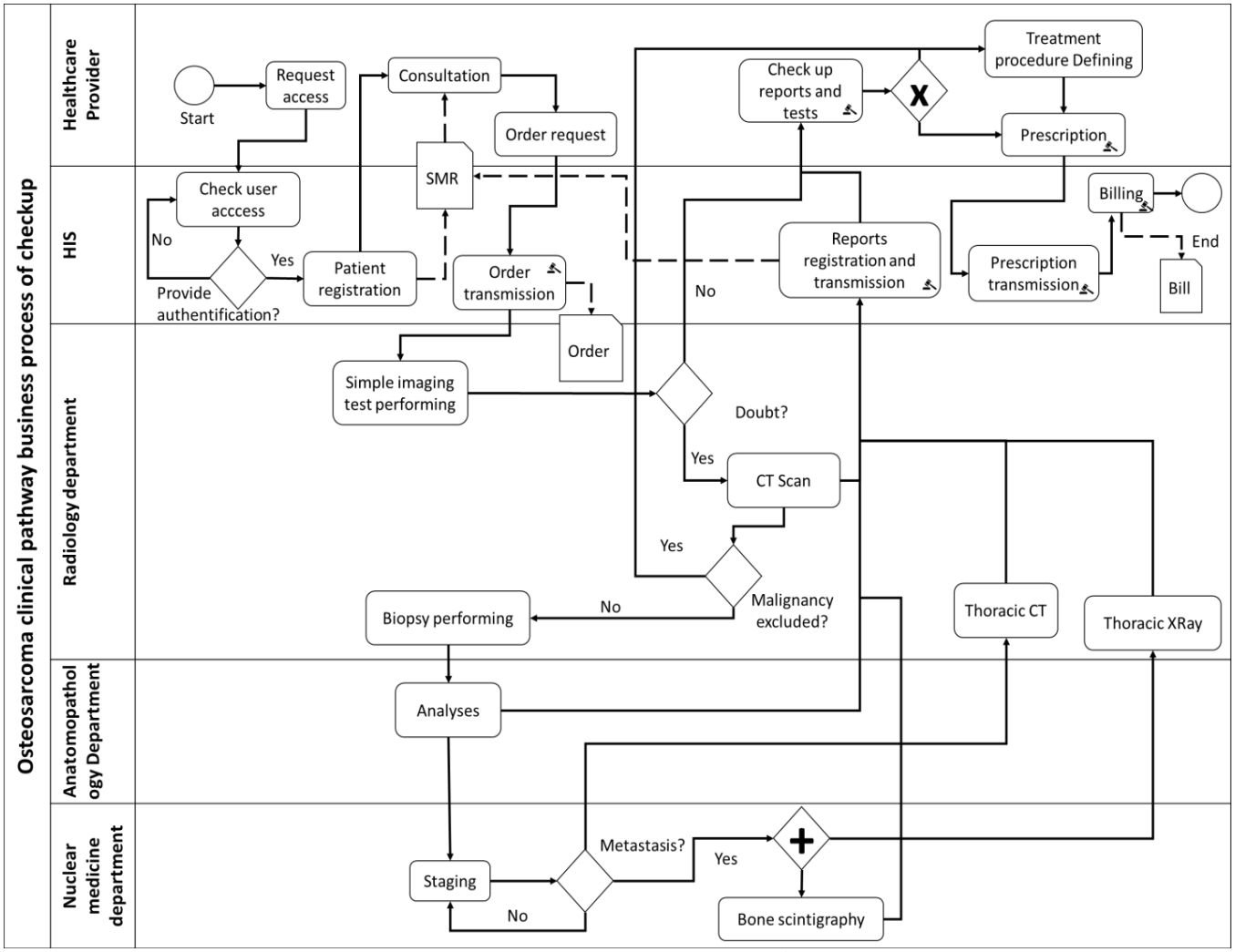


Figure 1. Osteosarcoma clinical pathway business process of checkup

2) Osteosarcoma treatment clinical pathway:

Based on tests' findings within the diagnosis phase clinical documents, the doctor defines the treatment phase according to systems review by biological tests, audiogram hearing tests and heart tests. By Luetke [23], during the treatment, medical tests and clinical examination are performed to evaluate its effectiveness as shown in Fig. 2. As the previous phase of the osteosarcoma clinical pathway, the treatment phase needs that many healthcare providers to get involved in it in order to be successfully fulfilled. Shared and transmitted medical documents in the course of this phase processes may be used for many purposes as the public healthcare and the public research. Besides, those documents are generally transmitted to the cancer registry in order to elaborate statistics. For that, the impact of the findings in medical documents (e.g., findings in chemotherapy, surgical operations and radiotherapy) on care quality improvement is very important. As a consequence, data discovery presents a crucial step to extract useful data for the medical data secondary use. This allows the sensitive data discovery which need to be protected in compliance with the HIPAA regulation.

3) Osteosarcoma follow up clinical pathway:

The last phase in the osteosarcoma clinical pathway is the following-up phase presented in Fig. 3. According to Paiolil [24], the doctor follows the patient health status by performing some tests and medical examination to check periodically the treatment effectiveness. Findings in medical documents managed in this phase may be used also for secondary use. Throughout the three phases of the osteosarcoma clinical pathway, diverse clinical documents are shared, transmitted and updated within the Shared Medical Record (SMR) ensuring the healthcare continuity.

Furthermore, clinical documents management enforced by Hospital Information Systems increases the risk of sensitive data disclosure or misuse. Despite the enhanced security measures while implementing network infrastructure and Hospital Information Systems, data hackers still find ways to penetrate networking systems and hack data. As a consequence, for a secondary use, the risk of sensitive data disclosure or misuse increases too. For that, we have resorted to sensitive data discovery managed throughout clinical pathways. This allows reinforcing its protection in compliance with the HIPAA regulation by predicting the risk rate of data disclosure that each type of medical document faces, then implementing computerized security methods in Hospital Information Systems since an early stage of their design for ensuring privacy enhancement.

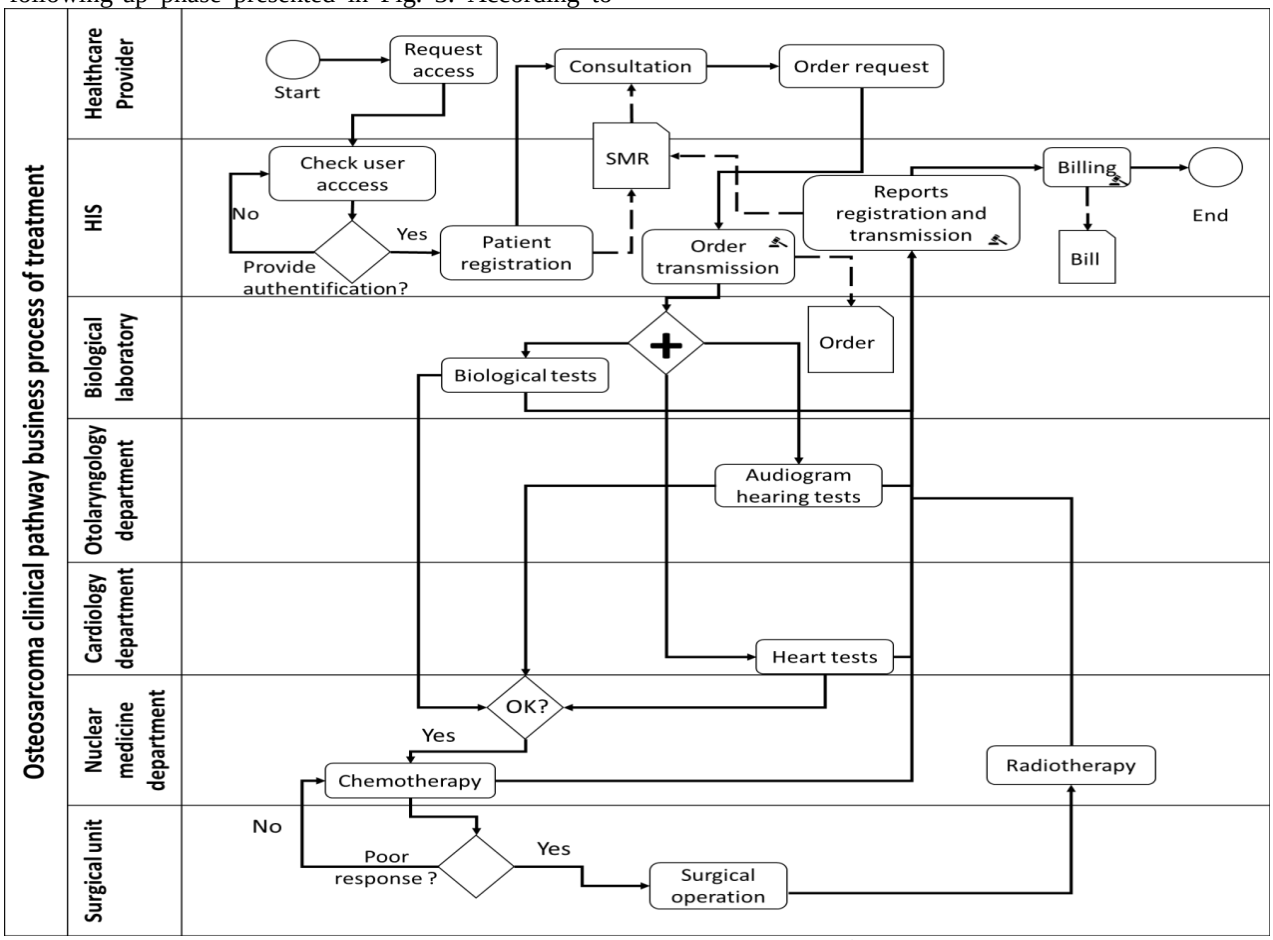


Figure 2. Osteosarcoma clinical pathway business process of treatment

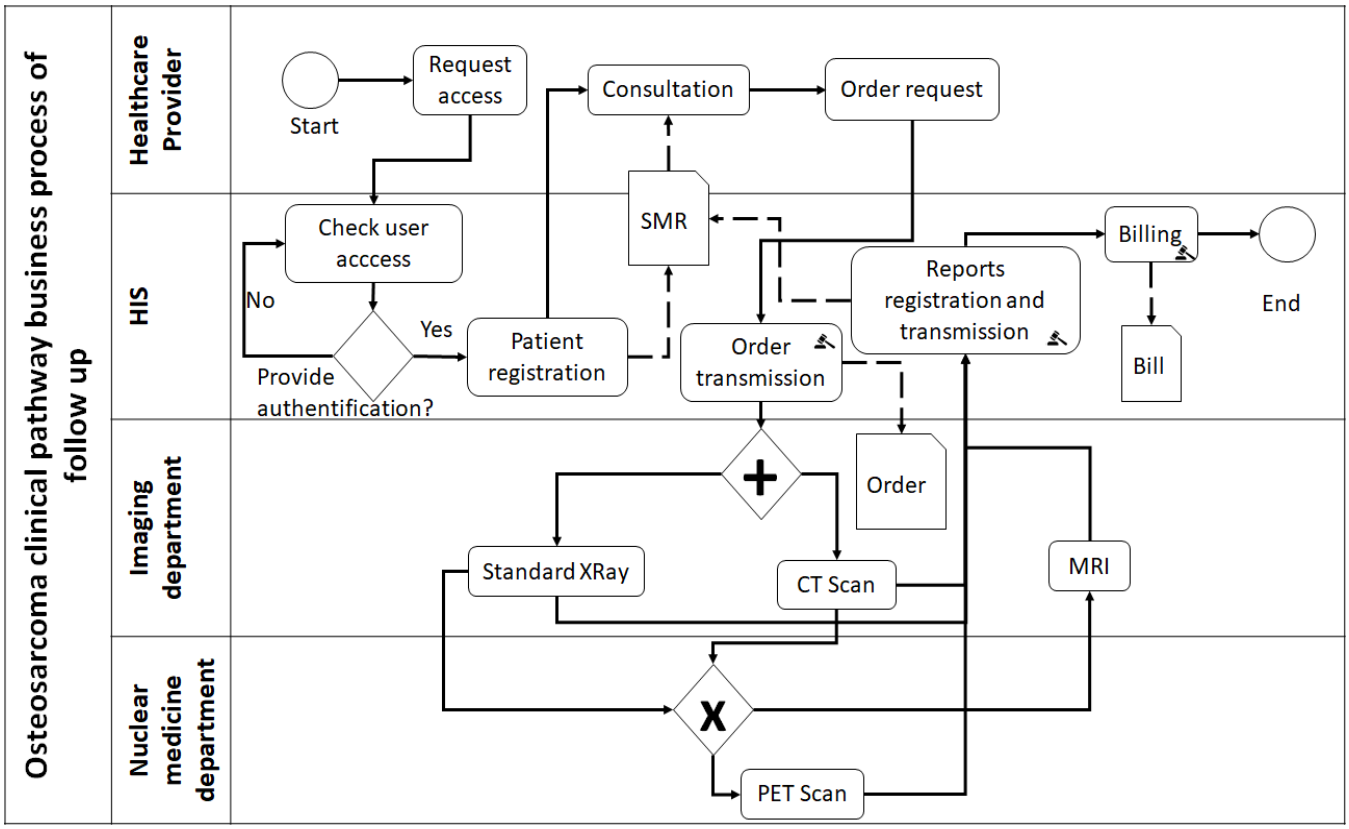


Figure 3. Osteosarcoma clinical pathway business process of follow up

IV. SHARED CLINICAL DOCUMENT ARCHITECTURE

A patient's Electronic Health Record (EHR) must contain all types of clinical documents including their medical history record, discharge summaries, typical paper charts, mental status examinations and other medical reports such as medical tests and operative reports.

Throughout a clinical business process, the EHR is transferred, updated and shared between healthcare professionals ensuring the continuity of care. Each clinical document included in the EHR contains medical data as it is required in the concerned healthcare establishment.

The general clinical document architecture is divided into documents, fragments and data. As shown in Fig. 4, clinical documents are composed of many fragments. They provide information about patients, procedures, practitioners, diagnosis, findings and appointments. The clinical shared documents' architecture model, illustrated in Fig. 4, could be adapted to another health care establishment, according to the used medical documents' structure in their boundaries. In each clinical document fragment, several medical data are found with specific properties which need the implementation of privacy by design approach. This is dedicated to the PHI use and disclosure within the HIS. The identification and demographic fragments in clinical documents include PHI. Its use should obey to the data protection law principles and privacy requirements ensuring the PHI privacy and the security of the medical data in use [26][27].

Based on data discovery using the HL7-CDA standard, the identification of sensitive data among the shared and the transmitted clinical documents facilitates the application of the required security measures. This could be applied by the use of security computerized methods (e.g., encryption, decryption, anonymization and pseudonymization) since an early stage of the design of the HIS.

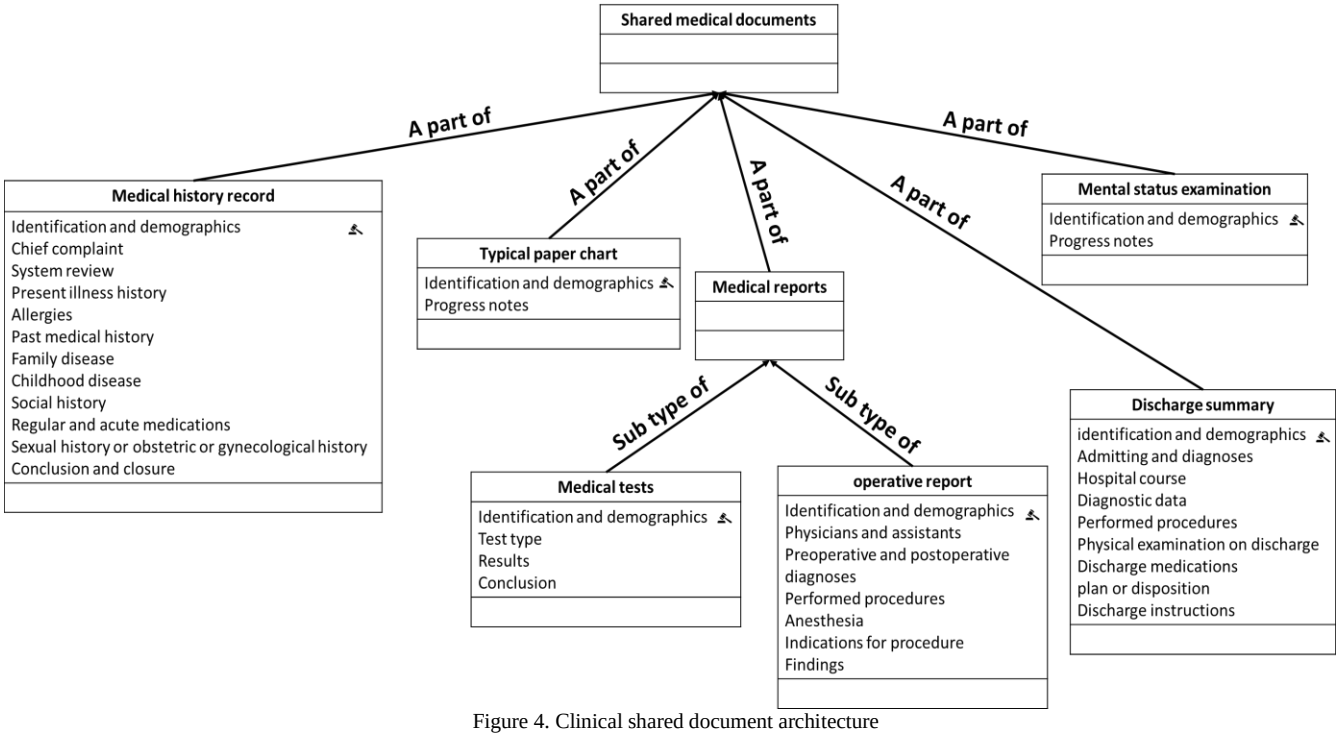


Figure 4. Clinical shared document architecture

V. HL7 CLINICAL DOCUMENTS ARCHITECTURE MAPPING

Data driven business process modelling enabled us to highlight shared clinical documents throughout different clinical workflows typically included in hospitals care pathways. Referring to the HL7-CDA standard documentation [18], we were able to investigate and define the shared clinical documents architecture as well as the included attributes. Thanks to the personal data protection principals, the PHI privacy requirements included in the HIPAA rules and the PHI defined within the HIPAA, we were able to discover sensitive data and highlight the PHI included in the shared clinical document.

The use of standardized architecture of clinical documents could facilitate distinguishing PHI from other included data. Furthermore, it facilitates the implementation of data protection methods while insuring a simplified clinical documents management and processing. The ontological representation of clinical documents' architecture allows capturing clinical information in the form of medical terminologies. Besides, it allows the classification of the clinical data into categories and the mapping each data category to a set of adequate data protection mechanisms as required by HIPAA Privacy Rule [3][4]. The use of ontologies helps us to deal with the nature of realism; it is the science or the study of being [28]. It allows the description of the included data in the clinical documents regarding to the human logic and philosophy to analyze the reality of things [29]. It generally provides a domain or a subdomain knowledge representation. A data model representation using ontologies generally provides a formal and conceptual model describing the logic of the data structure as well as the meaning of the data elements, which allow it to be understandable to both human and machine. Ontology-based models usually describe individuals (instances or objects), classes (concepts or types of objects), properties (attributes) and relations in a particular domain. Those relations describe ways in which classes and individuals could be related to one another [28].

Based on the CDA (Clinical Document Architecture) documentation of the HL7 standard, we mapped clinical documents to a set of characteristics aiming to provide a mean of classifying the included data in shared clinical documents. Each clinical document is characterized by a set of attributes providing information about the clinical document name, the category of each document (e.g., medical reports, medical record, etc.) and the clinical document architecture. This allows sensitive data discovery within shared clinical documents. After deep study of the general architecture of a clinical shared document, we could classify the included content into a set of metadata, data and values as shown in Table I and in Fig. 5.

The clinical document architecture is divided into two main parts: a header and a body. The first part of the document provides information about the document title, the document type, the document version, the participants, the organization, etc. As for the second part, it provides information about the clinical document content as the findings, reason for study, history, procedure context, study act and observation. So, we could map the clinical documents' data to attributes, as described in Table I. In this

part, we are taking an imaging report as an example to illustrate the clinical documents architecture according to the HL7-CDA standard [30]. The chosen document belongs to the category of *medical reports*. Particularly, it is a part of the *medical tests* as represented in Fig. 4. Most of the data values are associated with some standard variables part of CDA. However, many other variables are user defined to allow adaptation of the process model to specific organisational contexts. Some data could have suggested possible values as for the case of the *clinical document* type which can take a *CT Report* as a value for example. The included data in a clinical document has a hierarchical structure that specifies the recommended information for each shared document, participant, authorization, clinical statements and each section required to be present in this document according to its type. For example, for the *participants* section in the header of the imaging report, we found *record target* representing the patient's information, *author*, *data enterer*, *information recipient*, *legal authenticator* and *participant* as sub sections. For each sub section, a set of data values are required to be affected while the document processed and managed within the patient care pathway building the medical processes of the hospital information system. As it is illustrated in Fig. 5, the *record target* (patient) sub section affords information about the *patient role* which includes *address*, *phone number*, *patient name*, *birthdate* and his *gender*.

Data mapping and classification based on HL7-CDA standard presents the first step of our sensitive data discovery approach. As for the second step, it is based on the investigation of personal data protection principals and the extraction of the defined PHI within the HIPAA Privacy Rule as described in the following section. Matching findings of the first step with those of the second one allowed the sensitive data discovery within data driven clinical pathways.

TABLE I. CLINICAL DOCUMENT ARCHITECTURE DATA MAPPING

Characteristics	Attributes	Metadata	Data	Values
Clinical Document Name	Name	_	Value of the document name	Imaging report
Category	Category to which belong the clinical document	_	Value of the document category	Reports
Clinical Document Architecture	Header	Clinical Document	Title, date (effective time), version	Values are affected within the hospital information system
			Type	- Diagnostic Imaging Report - CT Report - MRI Report - Ultrasound Report - Nuclear Medicine Report - PET Scan Report - Cardiac Catheterization Report - Echocardiography Report - Colonoscopy Report - Endoscopy Report - Electrophysiology Report - Obstetrical Ultrasound Report
		Participants	Record target (patient), author, data enterer, information recipient, legal authenticator, participant	Values are affected within the hospital information system
		In fulfillment of	Order	
		Documentation of	Service event, physician reading study performer	
		Authorization	Information about the authorization	
		Related document	Parent document	
		Component of	Encompassing encounter	
	Body	Dicom Object Catalog	Study, series, SOP Instance UID	Values are affected within the hospital information system
		Findings	Sections (include paragraphs)	
		Optional sections	Reason for study, history, impression	
		Clinical statements	Procedure context, study act, text observation	

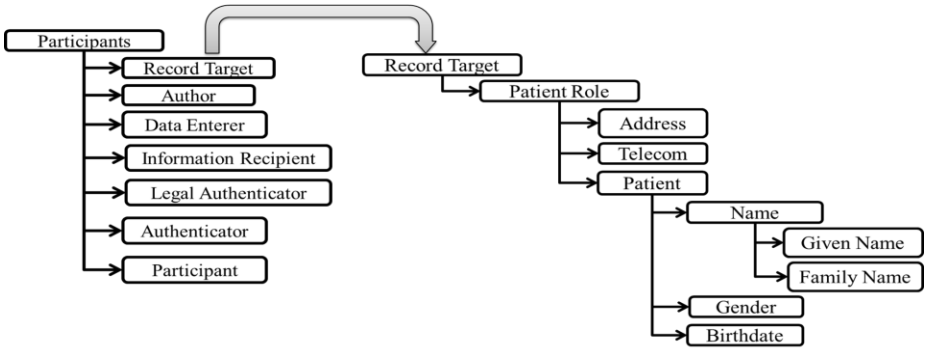


Figure 5. Example of the hierarchical structure of the participants section within an imaging report

VI. PRIVACY REQUIREMENTS FOR PHI PROTECTION

The use of clinical healthcare data is governed by many jurisdictions as it may present risks threatening a person's life and may affect both his/her privacy as well as his/her professional life. For this, clinical data usage must be set for data protection principles. In particular, the following eight principles should be respected as required in HIPAA regulation and the European directive:

1- Lawfulness, fairness and transparency: personal data should be processed lawfully, fairly in a transparent way.
2- Purpose limitation: personal data should be processed for specific purposes.
3- Data minimization: personal data should be adequate, relevant and limited to the precise purposes.
4- Accuracy: personal data should be kept up to date.
5- Storage limitation: personal data should be kept for no longer than the necessary period for the purposes for which those data are processed.
6- Rights: people have the right to access their data and give permission for other entities to use or disclose them.
7- Integrity and confidentiality: personal data should be processed in a secure way. They should be protected also against any unauthorized or unlawful processing, accidental loss, destruction or damage.
8- International transfers: personal data should not be transferred outside countries [26].

International law frameworks, such as European directive and HIPPA for personal data protection are based on the previous data protection principles. The present work is developed with regard to Protected Health Information within HIPAA regulation. The HIPAA Privacy Rule is published by the department of Health and Human Services (HHS) to ensure health information privacy. The privacy rule is applied to covered entities as health plans, healthcare clearinghouses and the healthcare providers. It defines a set of rules in order to protect sensitive health information with respect to its use and disclosure. Sensitive health information is known as Protected Health Information (PHI). They are individually identifiable health information related to the patient's past, present and future physical or mental health conditions, the healthcare provision to the individuals and the past, present or future healthcare provision to individuals [2]. The individually identifiable health information includes demographic data and many common identifiers. PHI usage and disclosure are permitted without the patient's informed consent for some purposes and situations as to the individual, the treatment, payment and healthcare operations, opportunity to agree or object, incidence to an otherwise permitted use and disclosure or public interest and benefit activities as well as a limited data set for research, public health or healthcare operations purposes or when it is required by law. As for the not permitted PHI usage and disclosures, an individual's written authorization (consent) must be obtained [3].

In addition to permitted PHI use and disclosure, prohibited ones are defined in Privacy Rules. For example, genetic information is considered as PHI and they shall not be used or disclosed for underwriting purposes as well as the psychotherapy notes. Furthermore, PHI may not be sold by covered entities. The PHI use and disclosure must be limited to the minimum necessary. However, PHI may be used to create a non-individually identifiable health information or a de-identified information [3][4].

The HIPAA Privacy Rule also defines a set of PHI de-identification requirements in order to use and disclose it without the patient's authorization. A covered entity may de-identify PHI by removing the eighteen identifiers specified in the following list as defined in HIPAA Privacy Rule:

1. Names.
2. Addresses with all geographic subdivisions smaller than a State.
3. Dates except year (birthdate, admission and discharge date, date of death).
4. Telephone numbers.
5. Fax numbers.
6. Email addresses.
7. Social security numbers.
8. Medical record numbers.
9. Health plan beneficiary numbers.

10. Account numbers.
11. Certificate/license numbers.
12. Vehicle identifiers, serial numbers and license plate numbers.
13. Device identifiers and serial numbers.
14. URLs (Web Universal Resource Locators.
15. IP (Internet Protocol) address numbers.
16. Biometric identifiers (finger and voice prints).
17. Full face photographic images and any comparable images.
18. Any unique identifying number characteristic or code [1].

For the above identified PHI usage and disclosure purposes, de-identification based on computerized methods is necessary to respect the PHI privacy and ensure its protection from any illegal use or other threatening risks. Furthermore, they allowed us to identify the PHI included in the shared clinical document with regard to a set of clinical document data attributes as described in Section V. Our sensitive data discovery is based on the defined PHI in HIPAA Privacy Rule. They could be defined as sensitive information while our clinical document architecture ontology development. This will facilitate clinical documents and data mapping to a recommended data protection techniques that ought to be HIPAA compliant. For that, we will first start by defining generic data protection techniques as required by HIPAA then we will highlight the applicable techniques according to the clinical document type and the contained sensitive data.

VII. RESULTS AND DISCUSSION

In this present work, we are interested in discovering sensitive data within the shared and transmitted medical document throughout clinical business processes. Therefore, we studied medical business processes in order to elaborate a data driven clinical pathway model, based on the BPMN language. Then, we mapped the clinical documents included data to a set of characteristics in order to provide a meaningful data classification. Based on the HIPAA Privacy Rule, we extracted personal data protection principals as well as the eighteen identifiers defined as Protected Health Information. The aim here is to ensure the respect of privacy requirements since early stages of HIS design. Then, we divided clinical data into categories and extracted PHI among them in the form of data model clinical pathway. After that, we defined both personal data protection principles and HIPAA privacy requirements for the specified PHI use and disclosure. Finally, all of the above listed objectives were validated through the modelling of osteosarcoma care pathway business process model chosen as a case study, mapping clinical data to a set of characteristics and discover sensitive data among them in order to facilitate our clinical data ontology development.

As for the completion of the modelling phase of osteosarcoma clinical pathway, we have modeled its complex care pathway which is divided into three phases: *check-up*, *treatment* and *follow-up*. This was done using the actual BPMN language simple patterns. Hence, personal data processing is integrated in the processes, particularly, in a legislation compliant manner which adds more trust to medical documents processing and sharing during the clinical process implementation. The shared clinical document data mapping to sets of attributes has allowed discovering sensitive data and classifying clinical data into formal concepts that are clearly structured and outlined. Data protection principals' definition according to the HIPAA Privacy Rule allowed the identification of sensitive clinical data or PHI included in shared clinical documents. This data ought to be protected using HIPAA compliant data protection techniques within the clinical document processing and management throughout clinical business processes and workflows. Data protection techniques will be defined according to the shared clinical documents' type and structure within the HL7-CDA standard and the HIPAA Privacy Rule compliance.

Many difficulties were encountered in clinical pathway modelling using BPMN due to the complexity and multidisciplinary aspect of medical procedures. This has led us to conclude the necessity of a more specialized care pathway modelling language. This has also highlighted the need for a new care pathway modelling and automation language that is sensitive-data driven and could integrate privacy requirements specification. Thus, a new extension of the BPMN modelling language is required.

VIII. CONCLUSION AND FUTURE WORK

Clinical pathways automation is highly required in standardized HIS. This is traditionally ensured by business process modelling. In this context, we developed a data driven clinical pathway business process model for osteosarcoma, as a case study. We used BPMN as clinical pathway business process modelling language. A shared clinical data model was elaborated further to the clinical business process model. Based on the HL7-CDA standard, we were able to discover clinical data and define its structure. Moreover, the mapping of the obtained data to a set of clinical document elements was necessary to specify the logic of the shared clinical data formal representation. This facilitates sensitive data discovery and PHI highlighting within the shared clinical documents referring to PHI defined in HIPAA Privacy Rule. This allows the application of personal data protection techniques in a more fluid compliance to HIPAA.

Since personal data management must obey to data protection law, we defined both personal data protection principles and HIPAA privacy requirements with relation to patients identifying medical documents, in terms of their both use and disclosure.

The adoption of privacy by design approach offers a better enforcement of privacy since an early stage of computer-based healthcare systems design. This allows an orthogonal integration of privacy obligations throughout the

clinical process. For this reason, we are working currently on sensitive data discovery within shared and transmitted clinical documents between healthcare providers. Mapping discovered data to a set of personal data protection requirements. Then, we identified the defined PHI within HIPAA Privacy Rule. This was very useful to discover sensitive data that need more protection than other shared data. This is crucial step to achieve our clinical documents' ontology development and apply HIPAA compliant measures. BPMN process modelling language need to be extended with privacy annotation features and additional patterns to allow modelling privacy specification as part of clinical processes and giving them more attention since an early stage of the HIS design. The definition of a common vocabulary qualifying clinical pathways specifications will more enforce the respect of privacy requirements. We believe clinical process modelling languages should be more adapted to a multidisciplinary clinical systems users' profile. The adoption of a variety of symbols and modelling patterns investigation is needed in order to better tailor to the requirements of the clinical community.

Sensitive data discovery based on data driven clinical business process models and PHI protection requirements within HIPAA compliance facilitated distinguishing sensitive data from other medical documents included data. This provides a mean of classifying the included data in shared clinical documents. Thus, it facilitates our clinical data ontology development which provides a simple way to associate the security level and risk to each sensitive data category.

REFERENCES

[1] I. Essefi, H. Boussi Rahmouni, and M. F. Ladeb, "Data Driven Medical Process Modelling for Privacy Protection in Care Pathways," The Seventh International Conference on Global Health Challenges (GLOBAL HEALTH 2018) IARIA, Nov. 2018, pp. 24-31, ISSN: 2308-4553, ISBN: 978-1-61208-682-8.

[2] U.S. Department of Health & Human Services *HHS: HIPAA Privacy Rule Summary*. [Retrieved: July, 2013]. http://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html.

[3] HIPAA Privacy Rule: Uses and disclosures of protected health information: General rules (§164.502). [Retrieved: January, 2013]. https://www.law.cornell.edu/cfr/text/45/164.502.

[4] HIPAA Privacy Rule: Uses and disclosures for which an authorization is required (§164.508). [Retrieved: January, 2013]. https://www.law.cornell.edu/cfr/text/45/164.508.

[5] Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). [Retrieved: March, 2017]. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2016.119.01.0001.01.ENG&toc=OJ%3AL%3A2016%3A119%3ATOC.

[6] L. Bouarfa and J. Dankelman, "Workflow mining and outlier detection from clinical activity logs," Journal of Biomedical Informatics, vol.45, pp.1185–1190, 2012, Dec. 2012.

[7] Z. Huang, X. Lu, H. Duan, and W. Fan, "Summarizing clinical pathways from event logs," Journal of Biomedical Informatics, vol.46, pp. 111 – 127, Feb. 2013.

[8] S. Adeyemi, E. Demir, and T. Chaussalet. "Towards an evidence-based decision making healthcare system management: Modelling patient pathways to improve clinical outcomes," Decision Support Systems, vol.55, pp. 117 – 125, Apr. 2013.

[9] R. Braun, M. Burwitz, H. Schlieter, and M. Benedict, "Clinical processes from various angles - amplifying bpmn for integrated hospital management," International Conference on Bioinformatics and Biomedicine (BIBM) IEEE, Nov 2015, pp. 837–845.

[10] M. Shitkova, V. Taratukhin, and J. Becker, "Towards a methodology and a tool for modeling clinical pathways," The 5th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2015) Procedia Computer Science 63, Sept. 2015, pp. 205 – 212, ISSN 1877-0509.

[11] A. A. Funkner, A. N. Yakovlev, and S V Kovalchuk, "Data-driven modeling of clinical pathways using electronic health records," Procedia computer science, vol. 121, pp. 835-842, Nov. 2017.

[12] E. Rolón, E. R. Aguilar, F. Garcia, F. Ruiz, M. Piattini, and L. Calahorra, "Process modeling of the health sector using BPMN: A case study," The First International Conference on Health Informatics-Diagnostic Pathology (HEALTHINF 2008) IARIA, Jan. 2008, pp. 173-178, ISBN: 978-989-8111-16-6.

[13] S. Ferrante, S. Bonacina, G. Pozzi, F. Pinciroli, and S. Marceglia, "A design methodology for medical processes," Applied Clinical Informatics, vol. 7, pp. 191-210, 2016, doi:10.4338/ACI-2015-08-RA-0111.

[14] F. Ruiz, F. Garcia, L. Calahorra, C. Llorente, L. Gonçalves, C. Daniel, and B. Blobel, "Business process modeling in healthcare," Stud Health Technol Inform, vol. 179, pp. 75-87, 2012.

[15] R. Braun, H. Schlieter, M Burwitz and W. Esswein, "Extending a Business Process Modeling Language for Domain-Specific Adaptation in Healthcare," The 12th International Conference on Wirtschaftsinformatik in Osnabrück-Smart Enterprise Engineering (WI 2015), Mar. 2015, pp. 468-481, ISSN: 0937-6429, ISBN: 978-3-00-049184-9.

[16] S. Bielack, D. Carrle, P. G. Casali, and ESMO Guidelines Working Group, "Osteosarcoma: ESMO clinical recommendations for diagnosis, treatment and follow-up," Annals of Oncology, vol. 20, pp. iv137-iv139, May. 2009, doi: 10.1093/annonc/mdp154.

[17] V. Augusto and X. Xie, "A modeling and simulation framework for health care systems," IEEE Transactions on Systems, Man, and Cybernetics Systems, vol. 44, pp. 30–46, 2014.

[18] HL7 standard. [Retrieved: 2019]. https://www.hl7.org/implement/standards/product_brief.cfm?product_id=7.

[19] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," Journal of biomedical informatics, vol. 61, 224-236, June 2016, doi: https://doi.org/10.1016/j.jbi.2016.04.007.

[20] N. Hashemian and S. S. R. Abidi, "Modeling clinical workflows using business process modeling notation. In Computer-Based Medical Systems," The 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2012), June 2012, pp. 1–4, ISNN: 1063-7125, ISBN: 978-1-4673-2051-1.

[21] K. D. Miller, R. L. Siegel, C. C. Lin, A. B. Mariotto, J. L Kramer, J. H. Rowland, K. D. Stein, R. Alteri, and A. Jemal, "Cancer treatment and survivorship statistic," CA: a cancer journal for clinicians, vol. 62, pp. 220-241, Jun. 2016, doi: 10.3322/caac.21149.

[22] M. S. Isakoff., S. S. Bielack, P. Meltzer, and R. Gorlick, "Osteosarcoma: current treatment and a collaborative pathway to success," Journal of clinical oncology, vol. 33, pp. 3029-3035, Sep. 2015, doi: 10.1200/JCO.2014.59.4895.

[23] A. Luetke, P. A. Meyers, I. Lewis, and H. Juergens, "Osteosarcoma treatment–where do we stand? A state of the art review," Cancer treatment reviews, vol. 40, pp. 523-532, May. 2014, doi: https://doi.org/10.1016/j.ctrv.2013.11.006.

[24] A. Paioli, M. Rocca, L. Cevolani, E. Rimondi, D. Vanel, E. Palmerini, M. Cesari, A. Longhi, A. M. Eraldo, E. Marchesi, P. Picci and S. Ferrari, "Osteosarcoma follow-up: chest X-ray or computed tomography?," Clinical sarcoma research, vol. 7, Feb. 2017, doi: 10.1186/s13569-017-0067-5.

[25] HIPAA Privacy Rule: [Online]. Available .from: https://www.hhs.gov/hipaa/for-professionals/privacy/index.html, last reviewed: 2015.

[26] M. C. Oetzel and S. Spiekermann, "A systematic methodology for privacy impact assessments: a design science approach," Eur. J. Inf. Syst., vol. 23, pp. 126–150, March. 2014, doi: https://doi.org/10.1057/ejis.2013.18.

[27] S. Rajamani, E. S. Chen, Y. Wang, and G. B. Melton, "Extending the HL7/LOINC Document Ontology Settings of Care," AMIA Annual Symposium Proceedings, Nov. 2001, pp. 994-1001, doi: https://doi.org/10.1016/j.cmpb.2015.09.020.

[28] C. Barki, S. Labidi, H. Boussi Rahmouni. "Ontology-driven generation of radiation protection procedures," World Academy of Science, Engineering and Technology, International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering, vol. 11, pp. 256-261, 2017.

[29] B. Smith, "Beyond concepts: ontology as reality representation," The third international conference on formal ontology in information systems (FOIS 2004) IOS Press, Nov. 2004, pp. 73-84.

[30] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, and A. Shabo, "HL7 Clinical Document Architecture, Release 2", Journal of the American Medical Informatics Association, vol. 13, pp. 30–39, Jan. 2006.

Strength Training – induced Left Ventricular Remodeling in Heart Failure Patients

Short Paper

Zahra Sadek, Said Ahmaidi
Laboratory EA 3300 (APERE) « Adaptations Physiologiques à l'Exercice et Réadaptation à l'Effort »
Picardie de Jule Vernes University
Amiens, France
Zahra.sadek@etud.u-picardie.fr
Said.ahmaidi@u-picardie.fr

Wissam Joumaa, Ali Salami
Laboratoire Rammal Hassan Rammal, équipe de recherche PhyToxE, Faculté des Sciences (section V),
Université libanaise, Nabatieh, Lebanon
wjoumaa@ul.edu.lb
a.salami@ul.edu.lb

Mahmoud Youness, Pierre Nassar, Righab Hamdan
Medical Research Center
Beirut Cardiac Institute
Beirut, Lebanon
Myouness55@yahoo.com
Nassar.pierre@hotmail.com
Mdrighabh@hotmail.com

Wiam Ramadan
Lebanese Institutes for Biomedical Research and Application (LIBRA)
International University of Beirut (BIU) and
Lebanese International University (LIU)
Beirut, Lebanon
wiam.ramadan@liu.edu.lb

***Abstract*— The hypothesis tested in chronic heart failure patients with inspiratory muscle weakness demonstrates that the Resistance Training results in further improvements compared to control in terms of dyspnea, cardiac, skeletal and inspiratory muscle function and quality of life. Twenty patients with ejection fraction ≤45% and inspiratory muscle weakness described by maximal inspiratory pressure <70% predicted, had undergone 3 exercise training sessions per week for 12 weeks. Patients were randomly allocated to one of two groups: RT group or control group. RT performed at 60% of 1 repetition maximum. Control group patients had no training at all. At the beginning and the end of the study, patients underwent pulmonary function test, respiratory muscle function test, echocardiography, exercise test, skeletal muscle function test, 6 minutes' walk test and were evaluated for their quality of life using the Minnesota living with heart failure questionnaire. RT showed significant positive effects on ejection fraction, exercise test variables, functional capacity, respiratory muscle function, skeletal muscle strength and endurance as well as dyspnea and quality of life. Resistance training was superior to control in all parameters assessed, and most importantly in exercise time (68% versus 27%, P=0.008), respiratory muscle function (Maximal inspiratory pressure, 60% versus 14%, P=0.001; Sustained maximal inspiratory pressure, 71% versus non-significant improvement in resistance training group, P=0.022) and skeletal muscle endurance. However, no significants improvements were detected in control group patients. Resistance training has been applied to improve respiratory muscle strength in chronic heart failure patients. In addition, the resistance training was safe and resulted in additional benefits in cardiac, respiratory muscle and skeletal muscle function compared to the control.**

Keywords- Chronic Heart Failure; Resistance Training; Functional capacity; Skeletal Muscle Function; Respiratory Muscle Function; Ejection Fraction.

I. INTRODUCTION

Exercise intolerance, dyspnea and fatigue are the main obstacles that Chronic Heart Failure (CHF) patients face during their daily life activities [1]. These phenomena might contribute to physical impairment and result in a poor Quality of Life (QoL) [2]. Such restrictions happen as a consequence of reduction in skeletal muscle mass and strength, which might be explained by both qualitative and quantitative abnormalities [3]. In fact, there are histological and biochemical derangements expressed respectively by altered fibers distribution and reduced oxidative enzyme activity in addition to muscle metabolism impairment, mitochondrial changes, inflammation and muscle atrophy [4]. Skeletal muscle abnormalities are mainly characterized by muscle fiber type switch, fiber atrophy and muscle wasting, as well as the reduction in mitochondrial volume density and mitochondrial aerobic capacity [5–7]. Similar alterations have been detected in the respiratory muscle, with the exception of fiber type switch. In the peripheral skeletal muscles, the switch appeared to be in favor of type Ⅱb fast twitch fibers with low aerobic capacity that gets quickly fatigued [5] while respiratory muscles had a switch towards type Ⅰ slow twitch fibers [5]. After a period of intense evaluation of the safety and effectiveness of exercise rehabilitation in CHF patients, exercise training was shown to be the cornerstone of cardiac rehabilitation programs.

Because muscle dysfunction represents a hallmark of heart failure, the emphasis was on Resistance Training (RT), in order to restore the normal muscle structure and function [8]. For many years, bed rest and limited physical activity were recommended for all stages and forms of heart failure; while exercise was not suggested [9]. Nowadays, however, the concept of cardiac rehabilitation, including exercise training, and specifically RT, is well spotlighted and highly

recommended recently, mainly because of studies that show its benefits in various outcomes [6]. In fact, application of such programs induces significant histological, metabolic and functional adaptations in skeletal muscles, thereby, improving patient's QoL. McKelvie was the first to demonstrate in 1995 that there are no significant differences between cycling and RT regarding left ventricle response in heart failure patients [10]. This study found similar results with those obtained from Meyer et al. where central hemodynamics was stable and well tolerated during resistance exercise [11]. In addition, from the research conducted by Grosse et al. who performed RT at 65% of 1 Repetition Maximum (RM), an increase of 80–102% of muscular endurance and 14.5% of VO2peak was reported [12]. Pu et al. who performed RT at 80% of 1RM, also showed a 43% and 13% increase of muscular strength and Six-Minute Walk Test (6MWT), respectively [13].

These findings are consistent with those of Levinger et al. who discovered 18% increase of muscular strength and 19% amelioration of VO2 peak after training patients at 40–80% of 1RM [14]. As shown above, the significant improvements in muscle strength and endurance, the adaptation of muscle mass and the increase in the QoL and functional capacity had been proved by many researchers after RT [11] [13] [14]. Each study performs RT according to specific characteristics such as intensity, duration, frequency, number of repetitions and sets of exercise. Overall, these features should be taken into consideration to avoid any cardiovascular stress, and thereby any harmful consequence. The aim of our study is to determine the effects of RT on skeletal and respiratory muscle function, functional capacity, cardiac function, dyspnea [15] and QoL in patients with CHF. The paper proceeds as follows: Section II describes the experimental design, data are analyzed in Section III, Section IV presents the discussion and, finally, Section V draws the conclusions.

II. METHODS

A randomized, single-blinded, parallel controlled study was performed in patients who were diagnosed with stable CHF and inspiratory muscle weakness (IMW) [maximal inspiratory pressure (MIP) <70% predicted] and recruited from Beirut Cardiac Institute. 53 patients were assessed for eligibility and 33 were excluded due to some limitations (figure 1). Eligible subjects [Ejection fraction (EF) ≤ 45%, NYHA class II or III, diagnosed with CHF for more than 6 months as long as there has been no admission to the hospital or any change in medications throughout the previous 3 months] were randomized to different exercising groups. Excluded subjects suffered from pulmonary limitation [Forced expiratory volume (FEV1) and/or vital capacity <60% of predicted], orthopedic or neurologic disease, had a history of significant cardiac arrhythmia, a history of myocardial infarction or a cardiac surgery over the past 6 months, non-echogenic, unstable, poorly controlled blood pressure and/or end-stage HF.

A written informed consent form was signed and obtained from all the participating subjects and all of them were receiving the same type of medication that included mainly beta blockers, Angiotensin converting enzyme inhibitors (ACE-I) or Angiotensin II receptor blockers (ARB) and Diuretics.

Furthermore, an approval for the experimental protocol was obtained from the Committee for Ethics in Research of Beirut Cardiac Institute. Since humans were included in this study, a certificate from the National Institutes of Health (NIH) for Protecting Human Research Participants (PHRP) was obtained. The patients received detailed information and gave written informed consent before their inclusion in the study.

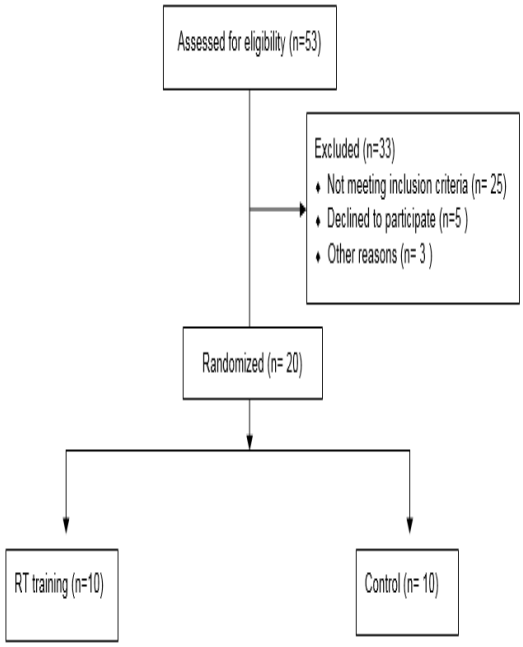


Figure 1. Flow chart of patient's recruitment.

20 subjects were block randomized with 1:1 to 2 groups as shown in the consort diagram in Figure 1. Randomisation was conducted by using a research randomiser web site (Scott Plous and Jeff Breil, Lancaster, Pennsylvania). The patients were divided, thereafter, into two different groups: controls (n=10) and resistance (n=10) for 12 weeks (3 times / week).

Demographic, clinical characteristics and drug therapy lists for all patients were collected at the beginning and the end of the study. Arterial blood oxygen saturation and heart rate (HR) were monitored using VIAMED pulse oximetry, and blood pressure (BP) was measured using a sphygmomanometer prior to the initiation and after the termination of each exercise session.

A. Cardiac Function

Echocardiography test: it is a test that makes an evaluation of the heart function possible. It creates images of the heart and helps in the analysis of different cardiac measures. Assessment of heart structure and function at rest was done using a Vivid S6 ultrasound probe (General electric healthcare GEMs supplies) where the patient was seated in a supine position. Two experienced cardiologists, blinded to subject's allocation and his/her study period, performed the test.

Measurements of ventricular volumes were assessed using the biplane method of disks (Biplane Simpson's method) from an apical 4 chamber view. Noting that this method is highly recommended for measuring the LVEF [16].

B. Exercise Test

Bruce treadmill protocol was used for the assessment of cardiovascular status during a progressive incremental exercise. Patients exercises on a treadmill with intensity and grade increasing progressively throughout the test following Bruce protocol. They also had ECG leads placed on their chests and a sphygmomanometer cuff for BP monitoring. Exercise time, grade, METs and BP were recorded during and after the test.

C. Lung Function

Pulmonary function was assessed using Digital Spirometer device. During this test, subjects were seated and instructed to inhale deeply through a mouthpiece well fitted between the lips and to exhale quickly and forcefully straight after. Air leakage was prevented by using a nose clip during the maneuver that was repeated three times to guarantee maximal performance.

The important parameters detected through this test were: forced expiratory volume in 1 s (FEV1), forced vital capacity (FVC) and FEV1 to FVC ratio.

D. Respiratory Muscle Strength

Respiratory pressure meter (MicroRPM carefusion), a portable hand-held device, was used to measure MIP and maximal expiratory pressure (MEP). The results of the test were displayed on a screen in cmH_2O. A PUMA PC Software that offers unique features could associate with the hand-held MicroRPM in order to improve its functionality. During this test, patients were instructed to tightly fit the mouthpiece attached to the instrument between both lips. Familiarization with the device and the ventilator operation was done before starting the test.

To measure MIP, the subject had to exhale to residual volume then perform Mueller maneuver characterized by a forceful inhalation against the instrument as long as possible.

To measure MEP, the subject must inhale to total lung capacity (TLC) and then exhale forcefully (Valsalva maneuver) over a minimum duration of 2 seconds. Both maneuvers were repeated at least 3 times in order to obtain the best values of MIP and MEP; noting that reported values were the maximum pressures sustained over a one second period.

E. Respiratory muscle endurance

To measure the endurance of the respiratory muscle, patients were instructed to breathe using Power Breathe set at 70% MIP/sustained maximal inspiratory pressure (SPI_{max}). The maximum time (in seconds) tolerated was recorded and meant to evaluate inspiratory muscle endurance, SPI_{max}/time. This test was repeated 3 times, with 1-2 minute rest intervals.

F. Leg Skeletal muscle function

Skeletal Muscle Strength Test: A handheld dynamometer (Lafayette instrument) was used to assess the skeletal muscle strength by measuring the force generated by the quadriceps. Because this device is portable and light, the examination procedure is made easier. Before the test, the subjects had been familiarized with the operative device and the intended protocol. Patients were placed in an upright seated position at a knee and hip angulation of 60° and 120° respectively. Verbal encouragement was used during the test period. Maximum voluntary isometric force (MVIF) in kg was measured; three sets of three repetitions were given to each patient to develop maximum force with a 20 second rest period between each set. MVIF was reported as the average of the nine values taken. Peak force, average force and peak time were recorded by the device and transmitted to the computer software.

Skeletal Muscle Endurance Test: 50% MVIF was used to assess skeletal muscle endurance. Patients were told to maintain 50% of the MVIF as long as possible until maximum tolerable elapsed time is recorded. The procedure was repeated three times separated by 5 minutes of rest. Endurance time was calculated as the average of the three sets. Verbal encouragement was given to patients before and during the test to achieve the best possible performance.

G. Physical Activity

Inclusion criteria for this study required that patients had a sedentary lifestyle to avoid bias since some patients might be following a specific activity program. Thus, in order to make sure that all the patients met the criteria, the short international physical activity questionnaire (IPAQ) [17] was used to assess the level of each patient's physical activity. The questionnaire was administered over the phone or during the first office meeting. It is made up of 7 questions the patients were asked to answer telling the time they spend doing an activity. The activities include those done at work, at home, yard work, exercise and others.

H. Functional Capacity

Functional capacity in CHF patients was predicted by using the 6 minute walk test (6MWT) [18], which is

evidenced to be reliable and reproducible. The test was performed in a 60-meter corridor under the supervision of a physical therapist. Initially, patients had been familiarized with the corridor space and the time needed to complete the test. BP, HR, and arterial oxygen saturation were measured before and after the test. A stopwatch was used to record the elapsed time as soon as the subject's began the test. Patients were free to decrease their speed and stop if necessary. The test was repeated three times with 5 minutes of rest between sets.

I. *Dyspnea*

During each training session and during the exercise test, patients were asked to evaluate their exertion level using Borg scale [15].

Borg scale provides information about perceived exertion but nothing on the level of difficulty of the exercise. The scale is a set of different intensity levels, where patients evaluate their exertion level using an attached visual numerical scale of 6 to 20 with 6 corresponding to no fatigue or other symptoms and 20 to maximum exertion. Borg scale is known for its sensitivity and reproducibility to be used as a subjective reference during submaximal workouts [19].

J. *Quality of Life*

QoL was assessed using Minnesota living with heart failure questionnaire (MLwHFQ), which was approved for the use in clinical practice and research [21]. MLFQ has been translated into and approved in Arabic.

This questionnaire is composed of 21 items used to qualify physical and emotional aspects. A grading scale that evaluates the degree of HF impact on the QoL, with 0 (none) to 5 (very much), was used. Low scores indicate that patients had mild problems associated with HF and high scores indicate a high impact of HF on QoL. Introducing the questionnaire to the patients was easy and questions were answered clearly.

The information collected thanks to MLHFQ provided an overview about the physical, social and mental status of patients before and after the interventional process.

K. *Training programs*

Patients were randomly assigned into two groups. The training regimens were performed 3 times a week for 12 weeks.

Control: The control group included 10 patients who did not exercise at all and were instructed to continue their normal life activities during the three months of trial.

Resistance Training: Ten patients were included in the RT group. The training protocol encompassed strength exercises that targeted the muscles of the quadriceps, the hamstrings and gluteus muscles of the lower extremities; Biceps, Triceps, Deltoid and Pectoralis of the upper extremities.

A total of seven strength exercises were used to train the upper body and six strength exercises to train the lower body limbs. Each workout was composed of 3 sets and each set of 10 repetitions, with at least 15 seconds of rest between sets. The patients started training at 60% of 1RM that was assessed and recalculated every two weeks. All the patients started with 3 sets of 5 repetitions each but after the third session, all of them had completed the whole 3 sets with 10 repetitions each. The total training time was 30 minutes.

L. *Statistical analysis*

All continuous variables are expressed as mean ± standard deviation of the mean (m ± SD). Baseline comparisons among the groups were performed using one way ANOVA-test for the normally distributed variables and Kruskal-Wallis test for the non-normally distributed variables.

The paired t-test and Wilcoxon signed-rank test were used to assess training induced changes (pre vs. post) within a particular group. The effect of intervention among the groups, the effect of time and the effect of group-by-time interactions were evaluated using repeated measures analysis of variance (RMANOVA).

Friedman test was used when data was not normally distributed, knowing that normality was tested using Kolmogorov-Smirnov test.

The inflation of type-I error due to multiple comparisons was controlled using Bonferroni rule. Statistical analyses were performed using SPSS software (version 20, SPSS Inc., Chicago, Illinois, U.S.A.).

III. RESULTS

A total of 53 patients were assessed for eligibility and 20 patients were recruited. After the first office meeting, 25 patients failed to reach randomization and eight others did not participate for the following reasons: long distance, family issues and other limitations.

No side effects have been recognized during the training period, and all patients in training groups seemed to be enthusiastic and interested. Compliance was monitored using a daily log booklet, and was excellent in the two groups.

Drugs were taken properly by all the patients according to their cardiac specialist instructions (ACE- or ARB, beta blockers, and diuretics).

Pharmacological therapy has not been changed over the 3-month training period. At baseline, there were no statistically significant differences among the two groups (Control and RT) with respect to demographic characteristics, drugs, Cardiac Function, Exercise Test, Pulmonary Function Test, Functional capacity, Qol Respiratory muscle function and leg skeletal muscle function (Table I).

Table I: Demographic and clinical characteristics of patient population

	Control	RT	P-value
Ages	52.6±11.2	55±6.7	0.440*
Females/Males	4/4	3/5	0.600*
BMI	31.3±3.4	33.3±4.9	0.514
Medication (%)			
ARB	10	50	0.254
B-blockers	33.3	33.3	0.994
Diuretics	38.1	28.6	0.773
ACE-I	36.4	31.8	0.899
EF (%)	36.1±8.2	36.6±2.6	0.729
LVEDD (mm)	58.3±6.6	64.6±5.2	0.069
LVESD (mm)	46±5.9	52±8.1	0.454*
LVEDV (ml)	130±38.9	135.9±38.6	0.797*
LVESV (ml)	93.3±30.3	98.8±30.3	0.781
Exercise time	475.5±144.7	421.5±124.4	0.532
METs	8.1±2.1	8.3±1.9	0.337*
MLWHF_score	29.9±10.6	34.1±13.6	0.968*
FVC	3.2±0.8	2.9±0.8	0.873*
FEV1	2.4±0.9	2.5±0.7	0.912
FEV1/FVC	74.9±11.7	86.6±11.6	0.090
6MWT (meters)	455.3±113.1	425.7±62.9	0.539
MEP (mmHg)	70.1±16.7	89.6±29.5	0.094
MIP (mmHg)	35.8±10.5	44.3±7.5	0.151
SPImax (seconds)	176±89.7	168.1±80.9	0.913
MVIF right (Kg)	18.0±4.7	19.7±4.1	0.578
Maintenance right (seconds)	105.0±30.3	90.6±24.9	0.536
Stage	2.8±0.9	2.6±0.9	0.921*
MVIF left (Kg)	18.1±3.7	21.2±2.6	0.152
Maintenance left (sec)	118±43.3	89.6±40.1	0.210
NYHA class	2±0.8	2.1±0.6	0.921*
Dyspnea	12.8±2.3	12.3±1.0	0.598*

Compared to control, the training group had 18% ($p<0.05$) and 11% ($p<0.05$) improvement in the right and left MVIF respectively, and a 24% improvement in the right quadriceps muscle endurance capacity (MT) ($p <0.05$).

Moreover, RT has shown a 27% improvement in exercise time ($p<0.01$), 24% in METS ($p<0.01$), and a 15% improvement in dyspnea sensation ($p<0.001$) (Figure 4). A significant increase in 6MWT distance (13%, $p<0.01$), a decrease in NYHA functional class (33%, $p<0.05$), and a decrease in MLWHFQ score (30%, $p<0.05$) were also noticed in figure 3. In addition, a little increase in (LVEF) was observed in RT (5%, $p<0.05$) versus no changes in the control group (Figure 2). Concerning respiratory muscle function, RT was able to improve MIP by 14% ($p<0.01$).

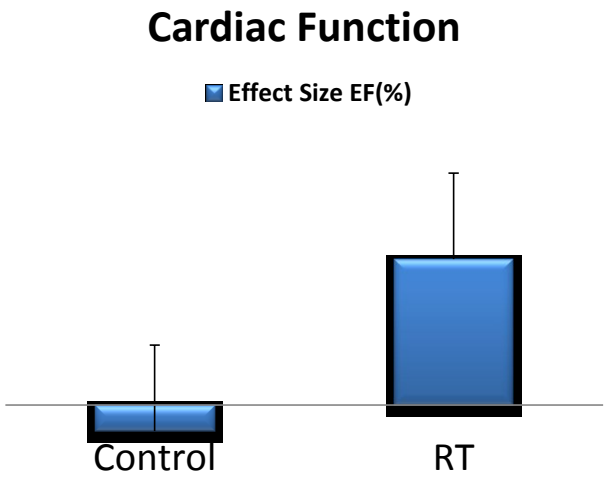


Figure 2. Effect size in cardiac function. Control: NS improvement in the cardiac function. RT: Significant improvements in EF (P<0.01). RT, Resistance Training; EF: Ejection Fraction.

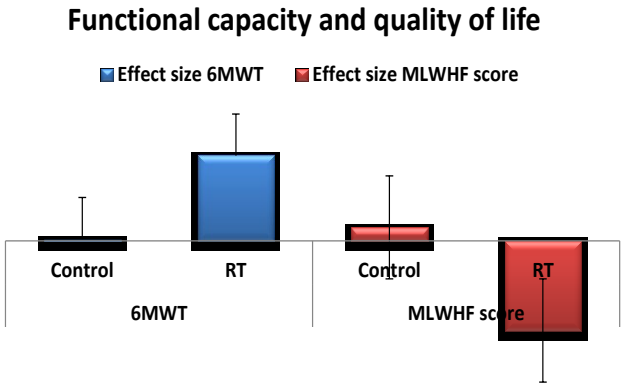


Figure 3. Effect size in Functional capacity (6MWT) and Quality of life (MLWHF). Control: No significant improvement in functional capacity and quality of life. RT: Significant improvements in 6MWT (13%, P=0.01), MLwHF score (30%, P=0.05).

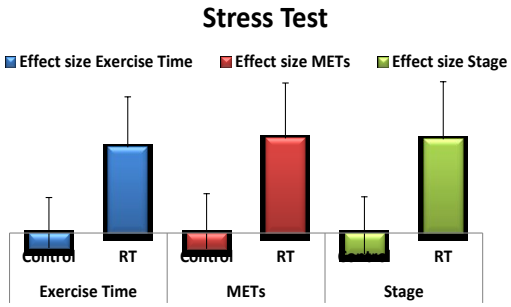


Figure 4. Effect size in Stress test. Control: No significant improvement in Mets (Metabolic Equivalent) and exercise time. RT: Significant improvements in Mets (24%, P=0.008), exercise time (27%, P=0.01).

IV. DISCUSSION

In this study, RT was able to bring out significant improvements in the skeletal muscle function, the cardiac function, NYHA functional class, and dyspnea as well as on functional capacity and QoL. Surprisingly, RT has also shown significant impact on respiratory muscle strength. Since the control group had no significant improvements in any of these parameters, we can confirm the effectiveness of the RT.

The benefits gained in skeletal muscle function affect positively the overall exercise performance and are closely related to beneficial adaptations in the muscle structure and function such as an increase in type 1 fiber, decrease in circulatory pro-inflammatory markers and a better muscle oxidative capacity [8]. In addition, such an intensive RT might induce an increase in the motor unit recruitment and so will impact on root mean square (RMS) value towards an upward trend.

In addition, we observed that the right quadriceps muscle is stronger than the left quadriceps muscle. These variations can be explained by the fact that maybe most of patients are right leg dominants.

Besides, as known, skeletal and respiratory muscle changes in heart failure are associated with biochemical and metabolic disorders. Thus, the fact of improvement in respiratory muscle function in our study may be attributed to the beneficial effects of RT in increasing mitochondrial enzymes and decreasing pro-inflammatory cytokines [4].

The skeletal muscle hypothesis confirms that impairments in the skeletal muscle not only alter the skeletal muscle by itself, but also contribute to further deteriorations and worsens the symptoms [21]. In the same manner as this hypothesis, RT also works not only by improving the skeletal muscle function, but also by enhancing the overall exercise performance and QoL, thus reducing hospitalizations as well as mortality rates. Therefore, regular RT programs are very efficient in counteracting these negative skeletal muscle abnormalities seen in CHF.

V. CONCLUSION

Considering our findings, we highly recommend the use of resistance training in CHF and IMW patients. This study had shown the crucial effect of RT, for it has improved the respiratory muscle strength. RT was safe and effective in improving skeletal muscle function, exercise performance, dyspnea and QoL in CHF patients, as well as in improving cardiac LVEF. Surprisingly in our study, RT improved respiratory muscle function in addition to improving cardiac left ventricular EF%.

We recommend that future studies investigate the usefulness of the electrical activity of the muscles known as Electromyography (EMG) in clinical diagnosis, in heart failure patients, in order to monitor the progression of skeletal muscle activity and function. Cardiac biomarkers could also be assessed in order to confirm the safety and effectiveness of interventions. Furthermore, studies at the cellular level could be a plus if added to the non-invasive measurements we used. Finally, biopsies of skeletal muscle might give a clearer insight about potential skeletal muscle adaptations after exercise conditioning.

It was not possible to assess aerobic capacity through measuring peak oxygen consumption (VO2max). However, we have used 6MWT and METs as prognostic outcomes. Both METs and 6MWT have been approved to be used instead of direct measurement of oxygen uptake [22], [23]. In addition, there is heterogeneity in training load, especially in exercise duration.

ACKNOWLEDGMENT

This research was supported by a fund from the National Council for Scientific Research (CNRS), Beirut, Lebanon; and was approved by the ethics committee of Beirut Cardiac Institute, Beirut, Lebanon.

REFERENCES

[1] Z. Sadek, S. Ahmaidi, M. Youness, W. Joumaa, C. Awada, and W. Ramadan, "Impact of Resistance Training in Patients with Chronic Heart Failure," The Seventh International Conference on Global Health Challenges, Athens, Greece, pp. 44–46, 2018.

[2] M. Grazzini, L. Stendardi, F. Gigliotti, and G. Scano, "Pathophysiology of exercise dyspnea in healthy subjects and in patients with chronic obstructive pulmonary disease (COPD)," *Respir. Med.*, vol. 99, no. 11, pp. 1403–1412, 2005.

[3] M. F. Piepoli and A. J. S. Coats, "The 'skeletal muscle hypothesis in heart failure' revised," *Eur. Heart J.*, vol. 34, no. 7, pp. 486–488, 2013.

[4] C. Zizola and Schulze P. Christian, "Metabolic and structural impairment of skeletal muscle in heart failure," vol. 6, no. 5, pp. 247–253, 2009.

[5] M. J. Sullivan, H. J. Green, and F. R. Cobb, "Altered skeletal muscle metabolic response to exercise in chronic heart failure. Relation to skeletal muscle aerobic enzyme activity," *Circulation*, vol. 84, no. 4, pp. 1597–1607, 1991.

[6] M. Schaufelberger, B. O. Eriksson, G. Grimby, P. Held, and K. Swedberg, "Skeletal muscle fiber composition and capillarization in patients with chronic heart failure: relation to exercise capacity and central hemodynamics," *J. Card. Fail.*, vol. 1, no. 4, pp. 267–72, 1995.

[7] H. Drexler, U. Riede, T. Munzel, H. Konig, E. Funke, and H. Just, "Alterations of skeletal muscle in chronic heart failure," *Circulation*, vol. 85, pp. 1751–1759, 1992.

[8] C. Delagardelle, P. Feiereisen, P. Autier, R. Shita, R. Krecke, and J. Beissel, "Strength/endurance training versus endurance training in congestive heart failure," *Med Sci Sport. Exerc*, vol. 34, no. 12, pp. 1868–1872, 2002.

[9] C. D. Mcdonald, G. E. Burch, and J. J. Walsh, "Prolonged Bed Rest in the Treatment of the Dilated Heart," *Circulation*, vol. 32, no. 5, pp. 852–856, 1965.

[10] S. M. Robert, Lm. Nei, T. Charles, B. Robert, and M. J.Duncan, "Comparison of hemodynamic responses to cycling and resistance exercise in congestive heart failure secondary to ischemic cardiomyopathy," *Am. J. Cardiol.*, vol. 76, no. 12, pp. 977–979, 1995.

[11] K. Meyer, K. Hajric, R.Westbrook, S. Haag Wildi, S. Holtkamp, R. Leyk, D. Schnellbacher, "Hemodynamic responses during leg press exercise in patients with chronic congestive heart failure," *Am. J. Cardiol.*, vol. 83, no. 11, pp. 1537–1543, 1999.

[12] T. Grosse, K. Kreulich, H. Naegele, "Peripheral muscular strength training in patients with severe heart failure," *Dtsch Z Sport.*, vol. 52, pp. 11–14, 2001.

[13] C. T. Pu, M T. Johnson, D E. Forman, J M. Hausdorff, R. Roubenoff, M. Foldvari, R A. Fielding, M A. Singh, "Randomized trial of progressive resistance training to counteract the myopathy of chronic heart failure," *J Appl Physiol*, vol. 90, no. 6, pp. 2341–2350, 2001.

[14] I. Levinger, R. Bronks, D. V. Cody, I. Linton, and A. Davie, "Resistance training for chronic heart failure patients on beta blocker medications," *Int. J. Cardiol.*, vol. 102, no. 3, pp. 493–499, 2005.

[15] I. Levinger, R. Bronks, D. V Cody, I. Linton, and A. Davie, "Perceived exertion as an exercise intensity indicator in chronic heart failure patients on Beta-blockers," *J. Sports Sci. Med.*, vol. 3, no. YISI 1, pp. 23–7, 2004.

[16] G. Magnusson, A. Gordon, L. Kaijser, C. Sylven, B. Isberg, J. Karpakka, B. Saltin, "High intensity knee extensor training, in patients with chronic heart failure. Major skeletal muscle improvement," *Eur Hear. J*, vol. 17, no. 7, pp. 1048–1055, 1996.

[17] P. Ponikowski, A.Voors, S. Anker, H. Bueno, J. Cleland, A. Coats, F. Andrew, G. Volkmar, R. José, V. Harjola, E.Jankowska, M. Jessup, C. Linde, P. Nihoyannopoulos, J. Parissis, B. Pieske, J. Riley, G. Rosano, L. Ruilope, F. Ruschitzka, F. Rutten, P. Van Der Meer, "2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure," *Eur. Heart J.*, vol. 37, no. 27, p. 2129–2200m, 2016.

[18] C. L. Craig Craig, A. Marshall, M. Sjöström, A. Bauman, M. Booth, B. Ainsworth, M. Pratt, U. Ekelund, A. Yngve, J. Sallis, P. Oja, "International physical activity questionnaire: 12-Country reliability and validity," *Med. Sci. Sports Exerc.*, vol. 35, no. 8, pp. 1381–1395, 2003.

[19] A. L. Beatty, N. B. Schiller, and M. A. Whooley, "Six-minute walk test as a prognostic tool in stable coronary heart disease: data from the Heart and Soul Study," *Arch Intern Med*, vol. 172, no. 14, pp. 1096–1102, 2012.

[20] S. Grant, T. Aitchison, E. Henderson, J. Christie, S. Zare, J. McMurray, H. Dargie,"A comparison of the reproducibility and the sensitivity to change of visual analogue scales, Borg scales, and likert scales in normal subjects during submaximal exercise," *Chest*, vol. 116, no. 5, pp. 1208–1217, 1999.

[21] T. S. Rector and J. N. Cohn, "Assessment of patient outcome with the Minnesota Living with Heart Failure questionnaire: Reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan," *Am. Heart J.*, vol. 124, no. 4, pp. 1017–1025, 1992.

[22] A. L. Clark, P. A. Poole-Wilson, and A. J. S. Coats, "Exercise limitation in chronic heart failure: Central role of the periphery," *J. Am. Coll. Cardiol.*, vol. 28, no. 5, pp. 1092–1102, 1996.

[23] J. R. Wicks, N. B. Oldridge, L. K. Nielsen, and C. E. Vickers, "HR Index—A Simple Method for the Prediction of Oxygen Uptake," *Med. Sci. Sports Exerc.*, pp. 2005–2012, 2011.

[24] S. Miyamoto**,** N. Nagaya, T. Satoh, K. S. Kyotani, F. Sakamaki, M. Fujita, N. Nakanishi, K. Miyatak, "Clinical Correlates and Prognostic Significance of Six-minute Walk Test in Patients with Primary Pulmonary Hypertension," *Am. J. Respir. Crit. Care Med.*, no. 3, 2012.

Wrist Actigraphy Analysis From Motionwatch8 Data

Gianluigi Delucca

Medricerca di Gianluigi Delucca
Rimini, Italy
email: ggdelucca@medricerca.it

Abstract—**This article expands the presentation made at SENSORDEVICES 2018. The author makes available to the public one year of wrist actimetry in real life, stored at one second epochs. The technical details of the system used (Motionwatch8) are described and they show why quantitative methodologies applied to actigraphy are tied to the system used for data acquisition. A first evaluation of that data calls for new ways for actimetry analysis and actigraphy display. Most of the values recorded are equal to zero and therefore the information they provide is a main target for actigraphy analysis. It is explained the difference of the quantity of zeros epochs when one second and one minute epochs are used. Series analysis and percent of zeros parameters are proposed and exemplified. In particular, the analysis of series is a new area of studies for actigraphy because it changes the meaning of the data recorded. The daily pattern of series is constant along the year and the importance of long zero-series is underlined: that small percent of lengths is not distributed randomly. The analysis of series also suggested the idea that the couples of series, not the single series, could be functionally similar and a actigraphy "pulse" analysis could be worthwhile to further evaluate. The percent of zeros parameter is applied to long and short time intervals and it could be a complementary tool for wake-sleep studies. The research described in this article is a ground breaking type of work that creates new questions more than providing answers, but it already offers new insights on actigraphy data.**

Keywords – Actigraphy; Actimetry; Motionwatch; Network physiology; Rhythms; Symbolic dynamics.

I. INTRODUCTION

This article expands the presentation made at SENSORDEVICES 2018, the Ninth International Conference on Sensor Device Technologies and Applications [1], adding details on the system used, the analysis suggested and some preliminary results.

Wrist actigraphy has a long history that starts in the 1950's and develops until today with research and clinical applications [2]. The terms “actimetry” and “actigraphy” are used as synonymous for motion activity measures where ”actimetry emphasizes the measurement aspect of the technique and actigraphy emphasizes the descriptive aspects of the technique” [3].

Clinical guidelines and researches suggest that wrist actimetry is particularly useful in the documentation of circadian rhythms, of sleep disorders, of treatment outcomes and as an adjunct to home monitoring of several pathologies. During the day, it is possible to quantify the physical exercise and (with calibration) recognize some type and intensity of the exercise. Reviews are available in several application areas [4]-[8].

Most published data use a one minute epoch, i.e., the system stores one piece of data each minute, and recordings are limited to a few days. The reasons for that selection of parameters are mainly practical, due to the characteristics of the available instrumentation and the logistic/organizational issues of the recordings, especially in real life. That selection of parameters sets limits on the possible evaluations of the data. Shorter epochs of actimetry could allow better correlations when the data is used for wake/sleep studies because the “gold standard” reference is the polysomnography (PSG), where data is presented in pages 30s long with signals in the range of 0-100 Hz. Without long continuous recordings, it is unknown if and how much Infradian rhythms, maybe spanning over years, have an impact on the evaluation of shorter recordings. The technology advances offer today the possibility to record more data and in an easier way, but still few long term recordings are described in literature [9]-[15] and only few groups explore shorter epochs [16]. None of those “extended” recordings are available to the public. To explore those two areas, the only option is to make longer recordings with shorter epochs and here we describe one of them that the author published on the National Sleep Research Resource (NSRR) platform [17]. That data expands the universe of wrist actigraphy both in length (from week to years) and in granularity (from minutes to seconds).

Section II describes the modality of the recording, the export of the data from the original format and the assembling of an activity file of one year.

Section III explains why each brand and each model of actigraphic systems needs its own computation of numerical data.

Section IV describes the main characteristic of the data, options to study it and the most important findings of those preliminary results.

Section V recaps the methodology issues and we conclude our work in Section VI.

Unless otherwise noted, the data used are those recorded by the author and available to the public [17]. Computations were made with the software Motionware (V.1.1.16) from CamNtech Ltd (Cambridge, UK), the manufacturer of the Motionwatch8 (MW8) activity monitor used for the recording [18], or by programs written by the author in Octave [19]. When data outside the published year is used, raw data are available from the author on request.

II. Materials and methods

The subject of the recording is the author: age 62 at the start of the recording, male, BMI= 26.3, no known major chronic pathologies. Data is collected using a MW8 on the non-dominant wrist. MW8 is a clinical system extensively used and it is a recognized reference in the field [20]. The monitor is set to store data in an epoch of one second in "normal" mode. That means that the intensity of the movement on the axis perpendicular to the surface of the unit, is measured by an accelerometer sampled at 50 Hz. Data is transformed in a single value of a custom unit (Counts) each second. The monitor acquires also a value of light intensity (Lux) each second. Details are described elsewhere by the manufacturer [21]. A maximum of 36 hours of activity and lights data is stored at 1s epoch in the unit and therefore there is the need for a data download every day. A diary is kept of major events (travels, flu, mismanagement, etc.). As usual in this kind of recordings, it is not possible to know the exact position of the unit on the wrist and if and when the photocell of the unit somehow was obscured (garments, tools, etc.).

The marker available on the logger is used to signal when the unit is not worn. Using the MotionWare software, the values inside the marked intervals are modified from 0 to "n/a". Then, the recordings are joined in files containing more or less 10 days (one million lines) and exported as .CSV files. Using a spreadsheet (OpenCalc – Apache OpenOffice – Apache Open Foundation), lines "n/a" are changed to "-1" and the data is divided in 9 columns: year, month, day, hour, minute, seconds, counts, lux integers, lux decimals. That format allows the files to be easily accepted by several software programming languages.

With programs we wrote in Octave, activity Counts are extracted, saved as .MAT files and joined in one year long file, for a total of 31.576.501 data lines. The time of the file is continuous, aligned to "summer" time (UTC+2: 27 March to 30 October 2016; 26 March to 29 October 2017) used in Italy in June, at the beginning of the recording. Evaluations and graphics of this article are computed from that file using programs we wrote in Octave.

III. Raw Data

This paper works over the data acquired with the activity monitor model Motionwatch8 (MW8) from CamNtech Ltd. It is not possible to directly compare the numbers out of that unit with other models or brands because each system has his own way of data acquisition, a well known situation not improved in the past 20 years [22], as described below.

A. Analog Band-pass.

At the end of last century few systems for wrist actigraphy were used, like the Geweiler from Sing Medical, Aktometer from ZAK, Actiwatch from Cambridge Neurotechnology, Motionlogger from Ambulatory Monitoring Instruments (AMI) and Colburn from IMSystems and they used different approaches. For instance, AMI offered a system with selection of ranges from 1 to 9 Hz and the Actiwatch used a 1-6 Hz bandpass. [23]. Observations from the group of Van Someren [24] started a change that moved Actiwatch to the bandwidth 3-11 Hz. Today, AMI product range is on a 2-3 Hz band-pass while CamNtech Ltd (heir of Cambridge Neurotechnology) stays on the 3-11 Hz. For other systems available in the market today, the band-pass seems an "implicit" value, not described in user manuals, white papers or "validation" studies. The issue seems not of interest both from a marketing and from a clinical point of view, meaning that the value is not even mentioned [6],[25].

B. Sampling.

Sampling rates of professional systems available in the market today, range from 30 to 256 Hz, the intensity range of the accelerometer is typically +/- 8G and quantisations used are from 8 to 12 bits. After sampling, the systems set a threshold in order to remove "noise" from different sources and below that level a 0 value will be stored.

C. Data reduction.

Data reduction is a common issue for many biological signals and especially for sleep, caused by the ability to collect much more data than humans like to manage. Old professionals that remember ink and paper polysomnography will probably also remember their old professor boasting that he was able to "read" recordings just looking on the side of the paper batch. It was a simple analog data reduction of more than 100 times because it was one line every two pages of 30 seconds data. For the MW8, the data reduction is made on line when only the maximum of the 50 values sampled each second is stored [26]. That decision implies the assumption that if there is some movement inside the second, then the integral of the movement will be a constant multiple of the peak value. In other words, that the integral of the movement will be similar to that of a rectangular form that changes only one side but not the other. An assumption without a published demonstration, but that is "good enough" from the manufacturer evaluation (personal communication). Those are the "raw data" available. A second step of the data reduction, possible on or off line, is the use of longer epochs, which is the sum of the one second values over an interval. If we use an epoch of 30 sec in the system settings, that will provide a number every 30 sec that become the raw data available out of the system.

With the situation described above, the algorithms used for data evaluation are tied to the individual system.

D. Measurement error.

The Micro electronic mechanical systems (MEMS) technology used today inside actigraphs does not suffer of the aging, lost calibration and inter system calibration differences as much as the old piezoelectric sensors. The calibration of systems (acceleration measurement) from the manufacturing is inside a +/- 5% for 1g (personal communication) and MW8 has an expected lifetime of 7 years without the need of any further calibration.

Unfortunately, the most important expected error is not coming from the instrument, but from its positioning. Even if the oval design of the MW8 is a nice improvement over the

squared models of the past, displacements along and around the wrist are still possible. Those displacements create a fluctuation from what would be the measurement if the system could be placed always in the same position for the full recording. We did not find any published paper regarding that issue. Most published articles that discuss and "validate" algorithms simply does not include the topic. So we set up a recording of two units on the same wrist for one month. It is not the best way to measure that error because the presence of a second unit by itself creates a bias. A better option would be to setup a grid of sensors around the wrist that covers all possible positions and then check the relationship among them. The development of such a grid would be useful also for other researches on the kinematics of the wrist. However, with the two recorders at least we get a dimension of the issue. Relative position, case and strap of the two instruments were randomly exchanged during the daily data download, trying to minimize the error sources different from the change of position on the wrist.

Table I shows the comparison of the results of the sleep analysis for the two units performed by the Motionware software out of 34 nights. Note that in order to compute the sleep analysis the Motionware software uses 30 sec epoch, summing up the 1 sec raw data.

Table I. Co-recording of two MW8 on the same wrist of 34 nights. Max positive, max negative and mean absolute differences of sleep analysis parameters between the two units.

	POSITIVE MAX DIFF	NEGATIVE MAX DIFF	TIPICAL VALUE	MEAN ABS DIFF
Lights out				
Fell asleep	00.08.00	-00.08.00		00.00.53
Woke up	00.13.00	-00.06.00		00.01.11
Got up				
Time in bed				
Assumed sleep	00.13.00	-00.07.00	06:31	00.01.44
Actual sleep time	00.12.00	-00.16.00	06:00	00.06.18
Actual sleep (%)	2,7	-3,7	92,1	1,3
Actual wake time	00.21.00	-00.12.00	00:31	00.06.21
Actual wake (%)	3,7	-2,7	7,9	1,3
Sleep efficiency (%)	2,5	-3,2	90,4	1,3
Sleep latency	00.08.00	-00.08.00	00:06	00.00.53
Sleep bouts	12,0	-7,0	30,0	2,9
Wake bouts	11,0	-7,0	30,0	3,0
Mean sleep bout	00.03.28	-00.02.41	00:12:01	00.00.50
Mean wake bout	00.00.44	-00.00.20	00:01:02	00.00.09
Immobile mins	12,5	-10,5	365,5	3,1
Immobile time (%)	1,3	-1,2	93,4	0,5
Mobile mins	7,5	-5,0	26,0	2,4
Mobile time (%)	1,2	-1,3	6,6	0,5
Immobile bouts	14,0	-5,0	38,0	2,8
Mean immobile bout	00.01.11	-00.01.58	00:09:37	00.00.29
Immobile bouts <=1min	4,0	-5,0	4,0	1,5
Immobile bouts <=1min (%)	5,9	-12,4	10,5	2,5
Total activity score	3198,0	-2344,0	2107,0	917,4
Mean activity /epoch	2,9	-2,3	2,7	1,0
Mean nonzero activity /epoch	24,2	-26,8	40,5	8,7
Fragmentation Index	5,7	-11,6	17,2	2,7
Threshold				
Rest per 24h (%)	3,7	-5,7	62,7	2,0
Average light (lux)	3,7	-0,4	0,2	0,5
Central Phase Measure (min)	7,5	-4,0	249,8	1,1

For each parameter, in Table I it is shown the maximum positive and negative difference, a typical value (in order to provide the reader with a relative dimension of the differences) and the mean of the absolute difference for all the nights.

Motionware software, offers also the statistics Non Parametric Circadian Rhythm Analysis (NPCRA) and 24h Average, as shown in Table II.

Table II. Co-recording of two MW8 on the same wrist of 38 days.

NPCRA Statistics:		
Start hour of analysis	11/08/2018 01:00	11/08/2018 01:00
Length of analysis	38 days, 9 hours	38 days, 9 hours
L5 Average	742	705
L5 Start Hour	01:00	01:00
M10 Average	8917	8878
M10 Start Hour	07:00	07:00
RA (Relative Amplitude)	846	853
IS (Interdaily Stability)	453	456
IV (Intra-daily Variability)	1115	1171
24h Average Statistics:		
Start time	00:45:00	00:46:00
End time	09:39:04	09:36:40
Start date	11/08/2018	11/08/2018
End date	18/09/2018	18/09/2018
Daytime Average Activity	2.6 MW counts	2.6 MW counts
Nighttime Average Activity	0.8 MW counts	0.7 MW counts
Day/Night Average Ratio	3.38	3.61
Fitted Cosine Peak	13:53:21	13:52:05

In this case all 38 days were used, including 4 days not fully co-recorded due to management mistakes, therefore adding more variability. Even so, the differences are minimal. If we perform those analyses over one week we find Table III, where there are some differences.

Table III. Co-recording of two MW8 on the same wrist of 8 days.

NPCRA Statistics:		
Start hour of analysis	10/08/2018 00:00	10/08/2018 00:00
Length of analysis	8 days, 0 hours	8 days, 0 hours
L5 Average	727	790
L5 Start Hour	00:00	00:00
M10 Average	7141	7405
M10 Start Hour	11:00	08:00
RA (Relative Amplitude)	815	807
IS (Interdaily Stability)	586	535
IV (Intra-daily Variability)	1342	1282
24h Average Statistics:		
Start time	00:00:00	00:00:00
End time	00:00:00	00:00:00
Start date	11/08/2018	11/08/2018
End date	18/08/2018	18/08/2018
Daytime Average Activity	2.3 MW counts	2.3 MW counts
Nighttime Average Activity	0.6 MW counts	0.6 MW counts
Day/Night Average Ratio	3.99	3.82
Fitted Cosine Peak	14:53:05	14:32:16

Those tables suggest that the error due to the movement of the instrument itself has a different importance for each parameter, depending on the length of the recording.

IV. RESULTS

Out of 31,576,501 samples, there are 24,936,212 zeros (78.97%), 1,034,485 n/a (3.28%) and 5,605,804 (17.75%) non-zero values. If we compare data stored at one second epoch with data stored at one minute epoch (most used epoch in published articles), we find a completely different

ratio between zero and non-zero samples. Over the 1440 minutes of a day, a typical result for one minute epoch would be 555 (39%) zero and 884 (61%) non-zero epochs while for the one second epoch, zeros are about 80 % of the total. How is that possible? The answer is that one minute epochs are computed as the arithmetical sum of 60 one second samples and most of those samples are zero. For instance, in the lower line of Figure 1, minutes 02.41, 42 and 43 would be counted as 3 active minutes using a one minute epoch, while there are only 5 seconds of detected movements.

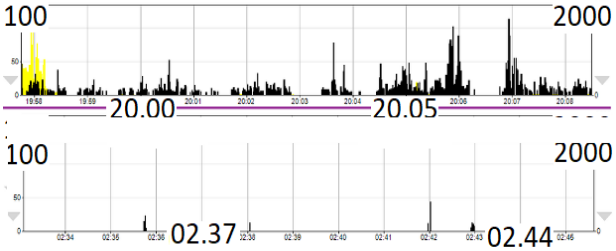


Figure 1. Examples of recordings at epochs of one second: day upper side, night lower. Actimetry in Counts (black) with range on the left, light in Lux (yellow) with range on the right.

Since the large part of the samples are zeros, we need to extract information also from those zero values. Here below we discuss some new possibilities.

A. *Series analysis*

One option for data analysis of zero values could be to show not the single epoch values, but the length of the succession of series of zero and non-zero values. For instance, the sequence ...,0,m,n,0,0,0,p,q,r,0.... will be described using positive and negative integers as …,2,-3, 3,...

In the recorded year, there is a sequence of 1.429.113 zero series and, of course, the same amount of non-zeros ones. The distribution of lengths of series at one second epoch in the year is shown in Figure 2, with non-zero series on the right and zero series on the left. The zero series can be long up to 7400 seconds, the non-zero series up to 623 seconds. There is a peak of several hundred thousand series around few short lengths.

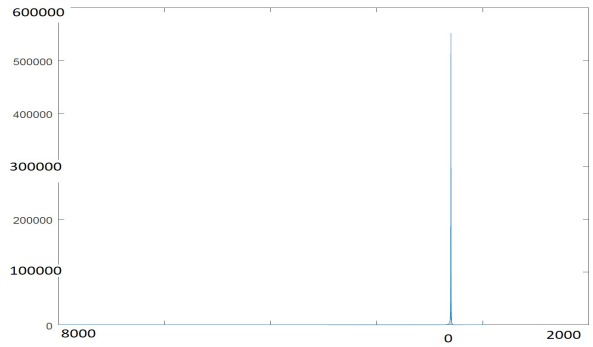


Figure 2. Full year. X axis: length of the series. Zeros series on the left, non-zero series on the right. Y axis: number of series.

If we zoom in Figure 2, we see (Figure 3) that, over one year, the number of zero and non-zero series longer than 10 seconds is a small percent of the total. But, that small percent of lengths are not distributed randomly. When we plot the series, with the length of non-zero series up and that of zero series down and the distance between the series equal to the length of the series, we get a graph as shown in Figure 4.

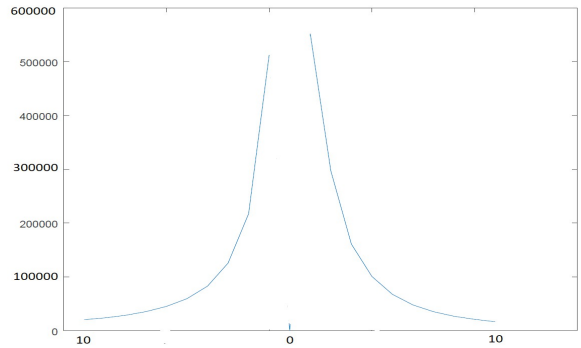


Figure 3. Full year - Zoom of Fig.4 X axis , zeros series on the left, non-zero series on the right Y axis– number of series.

The pattern is clearly bimodal and that daily profile is consistent over the year.

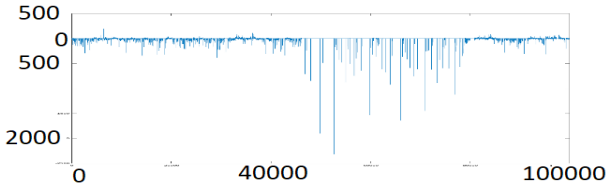


Figure 4. Day one, Start time 09:52:00. X axis: 100,000 seconds – Y-axis: length of the series in seconds: non-zero series up, zero series down.

We may then try studying those series as states of a system and search for models of their dynamics. For instance, we may evaluate one step from zero series or from non-zero series. If we plot the length of zero series as negative values and non-zero series that follows them as positive, as in Figure 5, we see that only short zero series can move to long non-zero ones.

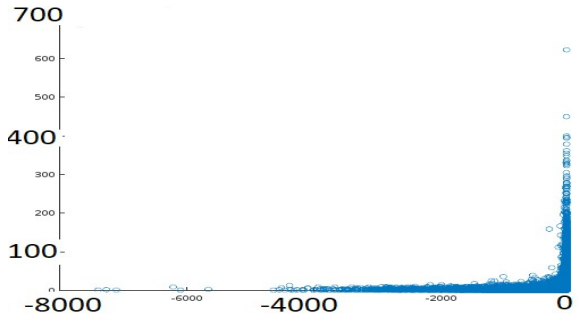


Figure 5. Full year. X axis: zeros series, length in seconds. Y axis : non-zero series that follow, in seconds.

In the same way, Figure 6 shows that only short non-zero series can move to long zero ones. This suggests, for instance, a model like the one in Figure 7 and the possibility to make statements on the data dynamics without any a priori hypothesis.

We know from Polysomnography (PSG) that the body activity changes during those long zero series because they cover different sleep stages and we then try to further investigate those long zero series. The year of data available on the NSRR platform is part of a larger project [27]. The recording started 6 months before of the published year and the internal report of those first six month is public [28].

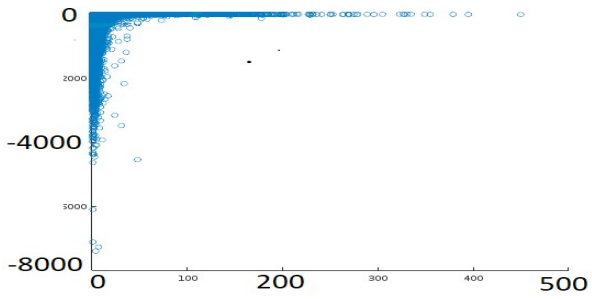


Figure 6. Full year. X axis: non-zeros series, length in seconds. Y axis: non-zero series that follow, length in seconds.

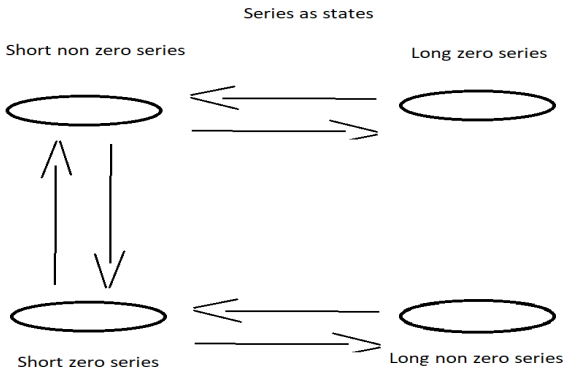


Figure 7. Model of series as states.

On the data recorded during the first month (27 days, one minute epoch) the minimum value of the longest zero succession of each night was 30 minutes, i.e., each night there is a zero-series long at least 30 minutes. Let's call D1 that group >=30 minutes. More than one series of that D1 group can be found in a night, up to four in that month. We may replicate the procedure (i.e., at least one value each night) below the D1 level and get two more groups: D2, from 16 to 29 minutes and D3, from 9 to 15 minutes.

If we add the hypothesis that all the D1 series are functionally similar, we may sum them and found a mean value of the total of a night of about 100 minutes. As described in that report [28], the analysis of the following 5 months (one minute epoch) afterwards demonstrated that there are nights without D1 series and the total D1 time in a night may vary from 0 to more than 230 minutes. More important, D1 series were possible only during sleep (personal communication). It was a first rough evaluation, but the hypothesis that long zero series are linked to physiology was worth to consider.

In a complex system like the human body, it is naive to think to find a fixed value for a threshold. It is more probable that the threshold, if it really exists. is fluctuating over time. For the moment, the fixed thresholds of D1 and D2 will provide a gross evaluation and if we evaluate D1 and D2 along the year we find a further support to the idea of two "types" of series hidden inside the D1-D2 ranges. We may sum day after day the length of the D2 zero series of the year and compare the real increment with a theoretic linear one. For a better display the linear sum is used as the zero axis and we get Figure 8. We keep the convention that zero series are a negative number. When the line is on the positive side, there is a cumulative shorter amount of D2 compared to the theoretic linear sum of the daily mean. When the line of the graph increases then that day there is less D2 then the theoretic linear mean.

Figure 8 shows a very small fluctuation similar to a bimonthly rhythm. The D2 daily mean is 6680 seconds and the mean length of a series is 1263 seconds, so about 5 series are needed each night.

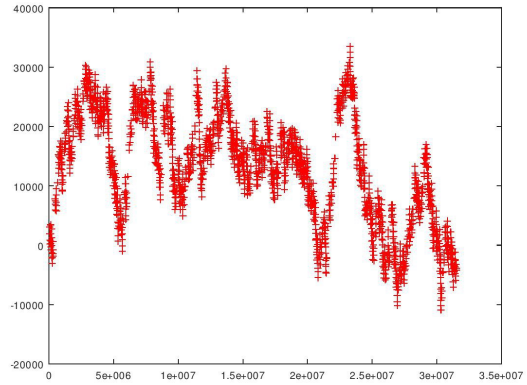


Figure 8. Difference of the D2 Zero Series sum less the linear sum of the mean along the year. X axis seconds from recording start, Y axis difference in seconds. Starting date 17 June 2016, near the summer solstice.

If we plot the values of the D1 series over the published year (Figure 9), we see that only few of them are longer than 5000 seconds (about 80 minutes), with a series mean of 2466 seconds. The daily mean is 4865 seconds, so one series is usually not enough to reach the daily mean.

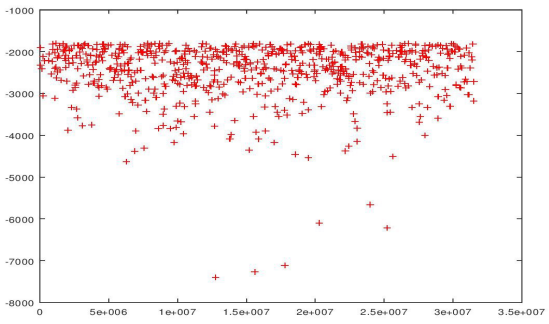


Figure 9. Zero series longer then 30 minutes (1800 seconds) along the year. X axis seconds from recording start, Y axis length in negative seconds.

When we perform for the D1 zero series group the same computation done for D2, we get Figure 10.

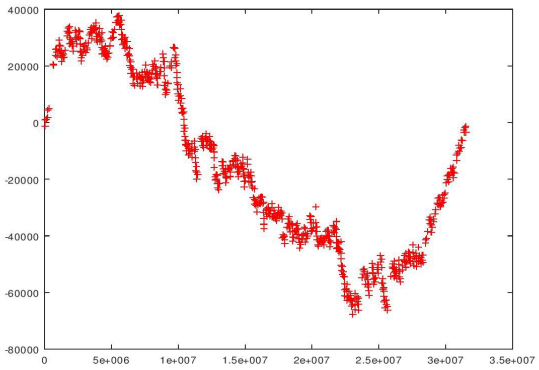


Figure 10. Difference of the D1 Zero Series sum less the sum of the theoretic daily mean along the year. X axis seconds from recording start, Y axis difference in seconds.

It is possible to note a small fluctuation similar to a circannual rhythm with peaks around middle August and middle April. The distance between the August - April peaks is about 80.000 sec (nearly 24 hours). That may seem a large gap, but that would mean that from August to April the mean daily length of D1 series is about 15 minutes longer than the April to August one, a small amount largely inside the day to day variability.

B. *Pulse analysis*

If we compute the basic model in Figure 7, we find that the central part of the short series is delimited by a threshold with daily values ranging from 20 to 40 seconds. That means that zero and non zero series long up to 20 to 40 seconds can freely move among them, but above those values that freedom is lost and longer intervals must be followed by series below that threshold. We may think that lost of freedom as the creation of couples of series, but it could also be that that "bond" is only more visible, i.e., that the "bond" that appears for long series exists, somehow, also for the sequences of short series. If we compute the distribution of couples of zero and non zero series, we find that the most important one is the (-1, 1) that covers 7% of the couples of the year. The couple (-1,+1) means that the accelerometer measures something that is shorter than one second and that is preceded by a pause longer than one second but shorter than 2 seconds. With the MW8 we cannot explore more in details that couple, due to the data processing characteristics explained in Section II. That couple is something that seems a pulse or a beat. We may suspect some kind of artefact, for instance from the heartbeat, but that would not explain the very long zero series, that would be randomly interrupted by an artefact.

The above was suggestive toward the analysis of actigraphic recordings as if those couples of series were "pulses". We may think of activity driven pulses (+/-) like "contraction and relax" or inactivity driven pulses (-/+) like "charge and discharge" and in the following we will consider the latter.

That "pulse" approach is new for actigraphy and in order to explore an uncharted area, an analogy could help. If we use the electrocardiography (EKG) analogy, then the actigram is like the electrocardiogram and the length of the "actigraphy pulses" (i.e., the length of one zero-series plus the following non zero one) is analogue to the R-R interval (where R is a point corresponding to the peak of the QRS complex of the EKG wave; and R-R is the interval between successive Rs). Figure 5 can then be "read" as the plot of the set of those pulses.

The range of R-R interval is about 200-2000 msec and the one of what we may call Wrist Actigraphy Pulse (WAP) is 2-8000 seconds. Those ranges barely touch each other and therefore R-R and WAP work in two different time scale. That would make those data complementary for the evaluation of lifestyles, and for sleep it has been already done [16].

So, it seems that WAP is a possible way to expand the study of zero and non zero series and that would put actigraphy in the same "basket" of other pulse analysis (HRV, Arterial pressure variability (APV), Oxygen pulse variability [29], etc.).

C. *Percent of zeros*

A more traditional approach, could be counting the zeros inside a defined interval.

1) *One minute.*

For instance, if we take one minute and we count the zeros, for each minute we get a number between 0 (maximal activity) and 60 (total immobility). The plot of the first day of the recording (1440 minutes) would be as in Figure 11.

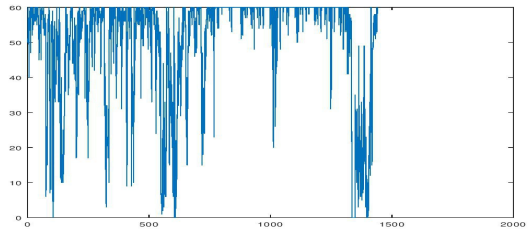


Figure 11. Number of Zeros -Day one Start time 09:52:00. X axis in minutes, Y axis number of zeros in a minute.

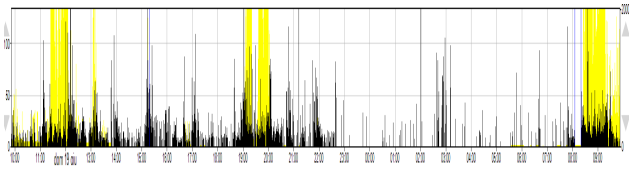


Figure 12. Day one at epochs of one second: Actimetry in Counts (black) with range on the left, light in Lux (yellow) with range on the right. Start time 09:52:00

The software Motionware provides a very similar value named "Immobile time Percent" inside the Assumed Sleep

Time (Table I), but accepting some very short activity [21], while in our case all values must be zero. We can compare Figure 11 with the standard Motionware activity plot in Counts as in Figure 12. With the zeros counting, it is easier to see that also during the day there are several minutes without any movements.

2) Three hours

If we divide the day in eight segments of three hours, we can count the number of zeros inside those intervals and get 8 numbers each day. It is also possible to express that number as a percent, i.e., percent of zeros (POZ) in 3 hours (POZ3h), and plot them in a circular way, instead of against a linear baseline. In that way, we get Figure 13 where the circumference is the year with the value 1 (total immobility) and the centre is the 0 (maximal activity) and there is a dot for each day.

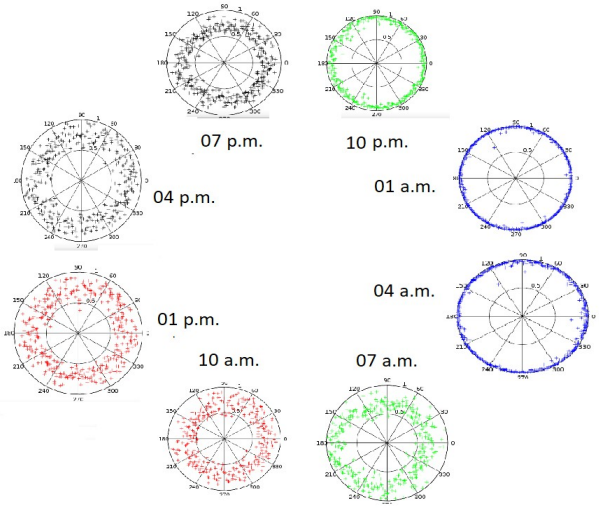


Figure 13. Full year. Zero values percent of 3 hours segments. The time is the start of the 3 hours segment. Circumference is one year, divided in months, 0 is 17 June 2016.

With that type of display, it is possible to see that a percent of zero values lower than 50% (internal circumference) is exceptional and that the circadian behaviour is consistent over the year in all segments.

3) Twelve hours

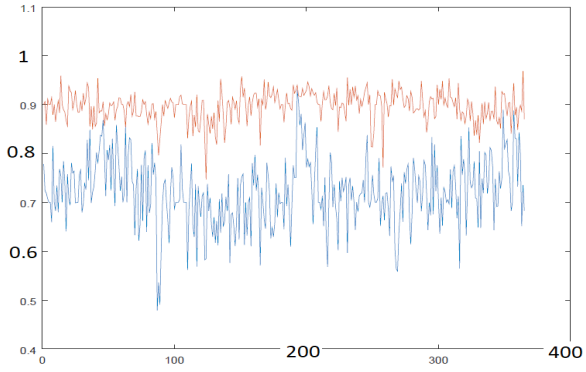


Figure 14. X axis: day of the year - Y axis: zeros percent in the time intervals 10am-11pm (blue) 11pm-10am (red).

If we divide the day in two pieces, like from 10 am to 11. pm and from 11 pm to 10 am, and compute for each day of the year the zero values percent of the two segments we can plot Figure 14. The two segments show a quite different behaviour along the year (few days with more than 5% of n/a data are removed for a better graph display).

The change of the clock to summertime (27 March to 30 October 2016; 26 March to 29 October 2017) may be responsible for some asymmetry in Figures 13 and 14 because solar time is from day 135 to day 285.

4) Segmentation

Motionware software offers the option to compute a 24 hours mean of the activity. If we average 3 months we see the pattern in Figure 15.

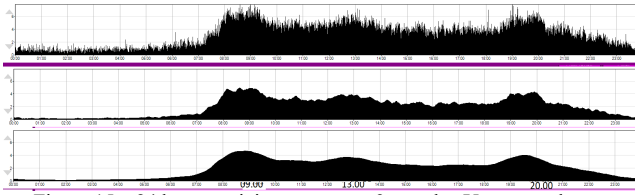


Figure 15 . 24 hours activity mean over 3 months. Upper graph no smooth, then 10 minutes smooth and one hour smooth. Start hour midnight. X axis 24 hours with hour divisions, Y axis Counts

Something similar would it be meaningful for POZ of 1 minute (POZ1)? If we plot the sum of POZ1 of 3 months for segments of 24 hours +/- 6 minutes, the pattern more visually similar to the one of a single day is exactly 24 hours (Figure 16).

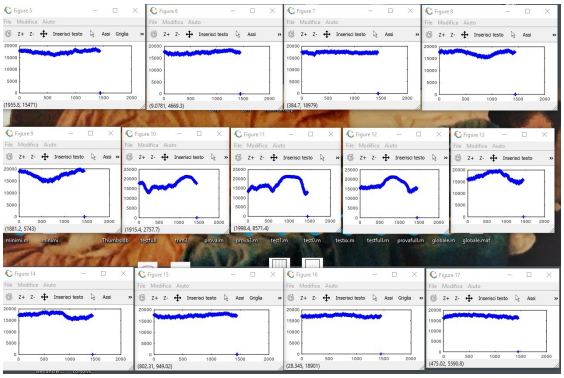


Figure 16. POZ1, 3 months sum of segments from 1434 to 1446 minutes. X axis minutes, Y axis POZ1 sum.

So, it seems that the segmentation of 24 hours is useful. We would get similar results from autocorrelation, but the methodology would require hypothesis on the data structure. The approach used, evaluating some segmentation and selecting the one more similar to the single day profile, is suitable for a machine learning algorithm and does not need hypothesis on the data structure. The development of that algorithm is a task for future work.

For the moment, let us accept the 24 hours segmentation. If we compute the sum of POZ1 on the 24h segmentation, we find that the parameter is sensible enough to show the summer hour shift (Figures 17 and 18).

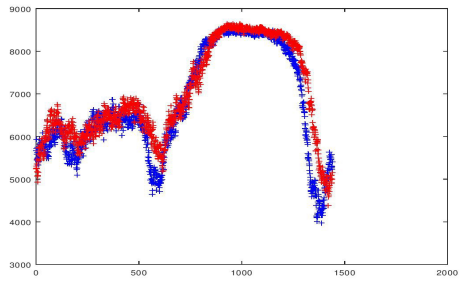


Figure 17. Sum of POZ of 1 minute (POZ1) over 5 months using a 24 hours segmentation (1440 minutes). Y axis POZ1 sum. Summer time 5 months in blue, standard time 5 months in red. Start time 09:52:00

It is possible to note that the profile in Figure 17 has two levels divided by two long transitions and therefore there is a large range of values that split the profile in two separated segments, as demonstrated in Figure 14. In Figure 13, those transitions are in the two segments 10 p.m.-01 a.m. and 07 a.m.-10 a.m.

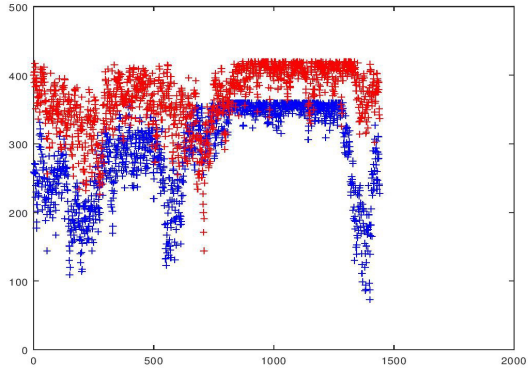


Figure 18. Sum of POZ1 over 1 week using a 24 hours segmentation (1440 minutes). Y axis POZ1 sum. Summer time in blue, standard time in red. Start time 09:52:00

The description of the plotted lifestyle of the subject would be: walk and breakfast (the first peak in Figure 15 and the last in Figure 17), office time, walk and lunch (peak), office time, walk and dinner (peak), after dinner until bedtime. Overall, it seems that the transitions from full wakefulness to deep sleep and back suggested by the activity levels, are long, taking more than one hour each. Those long transitions possibly explain why, in general, several different actigraphy systems and methodologies are able to recognise the wake and sleep circadian levels, because the selection of the threshold does not have to be precise. That will be true as long as the lifestyle pattern is "simple" as in this case and there are no important naps or long awakening at night.

It is worth to note that the use of POZ, like all the percent parameters, moves the scale of the parameter from the very high numerical difference of the Counts as in Figure 12, inside a recording and among recordings, to a fixed scale for all. The price to pay for that advantage is the losing of the information about the relative intensity of the movement. The "automatic" compression from the scale used for POZ1, makes the graph less spiky and usually there is no need for added operations in order to decide where the peaks are. We may then think of Figure 17 as the probability distribution of the movement of the non dominant wrist of the subject's lifestyle.

From co-recordings made after the published year, it seems possible to relate the slow down of activity after dinner with the dip of the core temperature [27]. So, it seems the POZ1 has the potential to become an useful complementary parameter for actimetry data analysis.

V. Discussion

A first step needed in any research is to understand the tools we use. Section II and Section III explain how the Motionwach8 works and why it differs from all other actigraphic system in the market.

Historically, the first objective of wrist actigraphy analysis was to classify sleep and wake states. An epoch of pre selected length (mostly 1 minute or 30 sec) is classified as Wake or Sleep, usually on the base of a threshold value, looking for a procedure that would provide parameters with good correlation to the PSG. The first set of parameters provided by the Motionware software (from "assumed sleep" to "mean wake bout" in Table I) are in that line of work. That evaluation is helpful as long as the time when sleep is expected (for the Motionware software the interval between the marks Light off and Got up) is well defined by external inputs [2].

The second attempt was to evaluate activity and immobility (in the case of Motionware software, the parameters from "Immobile mins" down, in Table I), using procedures more tied to the objective description of the data recorded. Traditionally, the data is analysed as a table of a dependent variable (activity Counts) and an independent one (time) with fitting methods like Cosinor [30]. That approach is quite difficult when 80% of values are zeros and push to focus on the study of those zero values. Over the years, there were studies on zero values, also called "immobility" [31]. It is one area of studies that is getting new attention [32] and the higher granularity of one second epochs has already shown some potential [16]. POZ may become another useful parameter because it is easy to compute and needs a small amount of memory, allowing long recordings for monitoring purposes. It offered already at least one result pointing out the permanent circadian distribution of the percent of zero values, as shown in Figure 13 and Figure 14.

However, the information provided by those analyses are always the same, no matter past, present and future computations and displays: we try to document and describe what a clinician would evaluate simply looking at the raw data in the standard printout of a week long recording i.e., if the daily pattern is regular or not, that there is more activity during the day than during the night (or similar consideration related to the working life), if the night time is without major interruption (Dr. Gioacchino Mennuni, personal communication). That description does not really needs a numerical evaluation, beside research studies.

In order to improve the situation, different approaches are needed. It is often overlooked the basic fact that any analysis method requires a model of the data we want to study. It is important then to underline that all the above historical analysis imply that any sample is considered equivalent to another, no matter when it is recorded, i.e., a movement of x Counts at noon has the same meaning of a movement of x Counts at midnight. And therefore making acceptable to compare the values measured in the two body states.

Sleep has always been related to the idea of "rest", connecting the seen "immobility" during sleep to the "immobility" of the wakefulness. After centuries, Electroencephalography (EEG) demonstrated the mistake showing that the highest level of immobility "REM muscle atonia" is a quite active state of the body. Nevertheless, the idea of sleep as "rest during an ongoing activity", is still entrenched (about someone asleep, we still say "he is resting") also in clinicians and researchers work, with methodologies that try to find something quantitatively less or more present instead of something qualitatively different. With the hypothesis that the series of zeros, not the individual sample, are functionally similar we enter a new area of studies. The epoch recorded has a different meaning depending when it is recorded, in this case, inside or not a series of a defined length.

The exploration of series and their dynamics is, as far as we know, a new area of research for actigraphy analysis. The series analysis demonstrates that, for some parameters, between wake and sleep there is not only a quantitative difference "more of the same immobility", but some qualitative one, with series that are possible only during well defined time frames.

We found that:

- the large part of an actigraphy recording is made out of short series;
- data is skewed, with a larger range for zero series than for non zero series;
- there is a threshold that limits the possible interaction among series based on their length and that divides them in "short" and "long" series;
- long zero series may be divided in groups (like D1, D2, D3);
- some of those groups are concentrated during the night with a circadian pattern constant over the year;
- the longest series of zero values are allowed only during sleep (personal communication).

The length of series seems not to be directly related to sleep stages and co-recording of polygraphic data is needed to explore that relationship.

The analysis of series suggested the idea that the couples of series, not the single series, could be functionally similar. The similarity of the couple of series with a pulse, opens the possibility to apply to the actigraphy data the large panoply of analysis already attempted for other "pulses". For instance, when labelling all non zero samples as "1", we create a binary coding suitable for Symbolic Dynamics. It is then intriguing to see a possible common framework for the analysis of different signals of the activity of the autonomous nervous system.

VI. Conclusion and future work

For the first time, we make available one year of wrist actimetry in real life, stored at one seconds epochs. The raw data described in this article are freely available to researchers on the National Sleep Research Resource (NSRR) platform .

Most of the activity values stored by the Motionwatch8 are equal to zero. Examples of ways to study those zeros and to display them are introduced.

The analysis of the published year is still limited:

- the issue of the "second door" of sleep has not been discussed;
- the search for a suitable approach to the study of actigraphy is still open and maybe this work will allow actigraphy to enter the network physiology field [33];
- it is unexplored if the intensity of the movement and the light recorded by the MW8 may add useful information.

The recording of the actigraphy inside the research is today more than three years long and there is a large area for further developments, for instance:

- all the issues related to the correlations with other biological signals are to be explored, starting from the difference between trunk and wrist actigraphy up to the correlations between actigraphy and PSG;
- of special cultural interest, at least for the author, would be to bring actigraphy, electromyography and eyelid closure under the same theoretical motor control framework.

The research described in this article is still a ground breaking type of work that creates more questions than providing answers, but new insights on possible analysis of actigraphic data are already obtained that stimulate to pursue the research. Hopefully, other researchers will look for new tools to analyse human motor activity and clarify the role of actigraphy in the personalised medicine.

VII. REFERENCES

[1] G. Delucca, "One Year of Wrist Actimetry at One Second Epoch in Real Life," The Ninth International Conference on Sensor Device Technologies and Applications (SENSORDEVICES 2018) IARIA, September 2018, pp. 185-189, ISSN: 2308-3514, ISBN: 978-1-61208-660-6.

[2] K.L Stone and S. Ancoli-Israel, "Actigraphy," In : Principles and Practice of Sleep Medicine, Edited by: M. Kryger, T. Roth and W. C. Dement. 6th edition Elsevier 2017 doi: 10.1016/B978-0-323-24288-2.00171-9.

[3] E.C. Winnebeck, D. Fischer. T. Leise and T. Roenneberg, "Dynamics and ultradian structure of human sleep in real life," Current Biology 28, pp.1-11. 2018.

[4] T. Morgenthaler et al., "Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: An Update for 2007," Standards of Practice Committee,

American Academy of Sleep Medicine, Sleep, vol. 30, No. 4, 2007 pp. 519-529.

[5] W. Pan, Y. Song, S. Kwak, S. Yoshida, and Y. Yamamoto, "Quantitative evaluation of the use of actigraphy for neurological and psychiatric disorders," Behavioural Neurology vol. 2014. Article ID 897282, 6 pages doi: 10.1155/2014/897282. Epub 2014 Aug 19.

[6] S. Ancoli-Israel et al., "The SBSM guide to actigraphy monitoring: clinical and research applications," Behavioral Sleep Medicine vol. 13, 2015, Suppl 1, pp. S4-S38, doi: 10.1080/15402002.2015.1046356.

[7] A. Castro, W.M. Anderson and R. Nakase-Richardson, " Actigraphy reference module in neuroscience and behavioural psychology," Encyclopedia of Sleep 2013, Editor-in-Chief C. A. Kushida, pp. 88–91, Reviewed 20 September 2016.

[8] G.J. Landry, R.S. Falck, M.W. Beets and T. Liu-Ambrose, "Measuring physical activity in older adults: calibrating cut-points for the MotionWatch8," Frontiers in Aging Neuroscience, vol. 7, 2015, p. 165, doi:10.3389/fnagi.2015.00165, ISSN=1663-4365.

[9] E. Werth et al., "Decline in long-term circadian rest-activity cycle organization in a patient with dementia," Journal of Geriatric Psychiatry and Neurology (JGP), 2002, vol. 15, pp.55-59, doi: 10.1177/089198870201500111.

[10] N. Lewis Miller and L.G. Shattuck, "Sleep patterns of young men and women enrolled at the United States military academy: Results from Year 1 of a 4-Year Longitudinal Study," Sleep, vol. 28, No. 7, 2005, pp.837–841.

[11] N. Lewis Miller, L.G. Shattuck and P. Matsangas, "Longitudinal study of sleep patterns of United States military academy cadets," Sleep, 2010 December 33(12), pp.1623–1631, doi:10.1093/sleep/33.12.1623.

[12] N. Lewis Shattuck and P. Matsangas, "A 6-month assessment of sleep during naval deployment: a case study of a commanding officer," Aerospace medicine and human performance, vol. 86, No. 5, May 2015, pp.481-485.

[13] C. Garbazza et al., "Non-24-hour sleep-wake disorder revisited – a case study," Front. Neurol. 7:17.2016, doi: 10.3389/fneur.2016.00017.

[14] J. Tippin, N. Aksan, J. Dawson, S.W. Anderson and M. Rizzo, "Sleep remains disturbed in patients with obstructive sleep apnea treated with positive airway pressure: a three-month cohort study using continuous actimetry," Sleep Medicine, vol. 24, August 2016, pp.24-31.

[15] A. Borbely, T. Rusterholz and P. Achermann, "Three decades of continuous motor activity recording:analysis of sleep duration," Poster P041 ESRS 2016 Bologna.

[16] A. Muzet et al., "Assessing sleep architecture and continuity measures through the analysis of heart rate and wrist movement recordings in healthy subjects: comparison with results based on polysomnography," Sleep Medicine, May 2016, vol. 21, pp.47–56.

[17] Data set available from: https://sleepdata.org/datasets/oya [accessed January 2019].

[18] CamNtech Ltd, Cambridge, UK, https://www.camntech.com [accessed January 2019].

[19] Octave.org - https://www.gnu.org/software/octave/about.html [accessed January 2019].

[20] Motionwatch bibliography, available from: https://www.camntech.com/products/motionwatch/mw8-biblio [accessed April 2019].

[21] CamNtech Ltd, Cambridge, UK, "Information Bulletin No.2," February 2016.

[22] S. Ancoli-Israel, et al., "The role of actigraphy in the study of sleep and circadian rhythms," American Academy of Sleep Medicine Review Paper, Sleep 2003, 26(3), pp.342-92.

[23] N. Stanley, "Actigraphy in Psychopharmacology," in Human Psychopharmacology, Measures and Methods, vol. 6, Edited by Hindmarch & Stonier John Wiley & Sons 1997.

[24] J.W. Van Someren, R.H. Lazeron, B.F. Vonk, M. Mirmiranam and D.F. Swaaba "Gravitational artefact in frequency spectra of movement acceleration: implications for actigraphy in young and elderly subjects," Journal of Neuroscience Methods, Volume 65, Issue 1, March 1996, pp. 55-62.

[25] Actigraphy Side-by-Side Comparison Guide (June 2016), Sleep review, available from: http://www.sleepreviewmag.com/2016/06/actigraphy-side-side-comparison-guide-june-2016/?ref=cl-title [accessed January 2019].

[26] CamNtech Ltd, Cambridge, UK, "Information Bulletin No.3," April 2016.

[27] G. Delucca. "Actigraphy analysis project," available from: . https://www.actigraphy.eu/actigraphy-analysis-project.html [accessed January 2019].

[28] G. Delucca. "Project Phase III report," available from: https://www.actigraphy.eu/actigraphy-analysis-project.html [accessed January 2019].

[29] S.A. Bhogal and A.R. Mani, "Pattern Analysis of Oxygen Saturation Variability in Healthy Individuals: Entropy of Pulse Oximetry Signals Carries Information about Mean Oxygen Saturation," Front. Physiol., 02 August 2017, doi:10.3389/fphys.2017.00555.

[30] F. Hallberg, F. Caradente, G. Cornelissen and G.S. Katinas, Glossary of Chronobiology, Chronobiologia 4 Suppl. 1, February 1977.

[31] A. Huub et al., "Circadian distribution of motor activity and immobility in narcolepsy: assessment with continuous motor activity monitoring," Psychophysiology, vol.32, issue3, May 1995, pp.286-291.

[32] T. Roenneberg et al., "Human activity and rest in situ," Methods of Enzymology, 2014, vol. 552, pp.257-283.

[33] P.Ch. Ivanov, K. Liu and R. Bartsch, "Focus on the emerging new fields of network physiology and network medicine," New Journal of Physics, 18(10):100201, October 2016, doi: 10.1088/1367-2630/18/10/100201.

Enforcing Genetic Consent and Restrictions through a Privacy-Focused Ontology

Michael Reep, Bo Yu, Duminda Wijesekera,
Department of Computer Science
George Mason University
Fairfax, VA, USA
e-mail: mreep@gmu.edu, byu3@gmu.edu, dwijesek@gmu.edu

Paulo Costa
Dept. of Systems Engineering and Operations Research
George Mason University
Fairfax, VA, USA
e-mail: pcosta@gmu.edu

Abstract— **The use of genetic information has greatly expanded from the original focus of providing actionable data to health care providers and researchers for diagnostic and research purposes. Potential uses of this information encompass the insurance industry, employment, and law enforcement plus the more recent development of Direct-to-Consumer (DTC) tests for genealogical research. Federal and State Laws have been developed in the United States to improve privacy protections and prevent the misuse of genetic data. However, there is a wide variety of laws, regulations and restrictions governing the release criteria, level of protection required, and specificity in permitted use. The attribute-focused component of these laws matches information regarding the requester, genetic contributor with the purpose and data being released to come up with an access decision. While the attribute-based portion is easily implemented, there are numerous aspects in the laws and regulations that require more complex decision making, dictate further post-release restrictions, and specific directives for consents. A rule-base specification of these complexities can be used as a policy language to enforce data releases from electronic health records and gene pools. Our previous work developed the attribute focused aspect of the ontology along with a workflow-based prototype. The final refinements to the ontology address the more complex requirements for consent, situational validations that must be confirmed, restrictions that must be enforced after data release, actions for data protection, retention and destruction by the recipient, and informing the genetic data recipients of potential penalties for violating these restrictions. Overall this framework provides the foundation for bolstering privacy protections, enforcing the laws and regulations, and preventing the unlawful disclosures of genetic information.**

Keywords- Genetic Privacy; Electronic Medical Records; Ontology; Health Care; Genomic Medicine, Informed Consent.

I. INTRODUCTION

Numerous issues must be considered when providing comprehensive protections for genetic information. Our on-going efforts focus on providing a comprehensive framework for consistently and vigorously enforcing the laws, policies and regulations related to protecting this vital information and implementing appropriate patient consents [1]. Patients are less likely to share medical data if there is a concern about privacy, so consents are necessary to help allay these concerns [2]. Privacy concerns have been heightened as Electronic Health Records (EHRs) have become widespread and therefore most of the information are on line, so can be accessed either legally or by other means – and consequently, ensuring privacy has increased in importance [3], [4]. There are demonstrable benefits to using genetic information as genetic studies map genotypic and phenotypic data directly to diseases, allowing for preventive and early interventional care to reduce morbidity, quality of life and treatment costs [5], [6]. In addition, studies in pharmacogenomics work to use genetic information in improving the effectiveness of drugs and reduce toxicity [7]. These benefits have to be balanced against inherent unusual characteristics of genetic information that can identify a patient and his/her genetic relatives, therefore placing any of them at risk of negative consequences, such as discrimination [8], [9]. Patient concerns extend beyond the inappropriate release for insurance and law enforcement to include access within the healthcare community [10]. Consequently, laws impose penalties if genetic data is inappropriately released. Studies have also shown that de-identification of genetic material may be insufficient to protect patient privacy [11], [12].

In the United States, overall health privacy was addressed by the Health Insurance Portability and Accountability Act (HIPAA) of 1996, which was implemented to improve the efficiency and effectiveness of the US healthcare system. HIPAA was followed by the Privacy Rule in 2000 to address three covered entities: health plans, health care clearinghouses, and certain health care providers [13]. The Genetic Information Nondiscrimination Act of 2008 (GINA) was passed to protect individuals from discrimination in employment and insurance based on genetic information [14]. Furthermore, almost every state and the District of Columbia have laws that specifically address genetic protections to some degree. However, even when patients are specifically provided information on GINA as part of pharmacogenomic testing, the subjects still report having little understanding of the act or privacy protections [15]. Health Information Exchanges and direct sharing between health care providers are still subject to the applicable state laws even for interstate data transfers [16]. This paper further develops an ontology that provides the syntactical elements (i.e., entities and their relationships) sufficient to specify applicable legislation and regulations in the forms of a machine enforceable structured rule-base.

In our previous work, we developed a prototype that uses a medical workflow system for an EHR to enforce Federal and

State laws in addition to organizational policies [17], [18]. Workflows provided the mechanism to gather the necessary information within the context of request to share genetic information in an EHR. We prepared an initial genetic privacy ontology and sample rules to enforce laws in selected states to validate our approach. We then extended the genetic privacy ontology, based directly on relevant Federal and state laws, to focus on the attribute-focused components that generate the initial access decision [18]. This paper provides further expansion and refinement of the more complex legal requirements for ensuring the appropriate consent is obtained prior to information release, validating pre-conditions for release have been addressed, and establishing the post-release protection mechanisms.

Our next step is developing this comprehensive genetic privacy ontology based directly on relevant Federal and State laws. Following this Introduction, Section II addresses related work; Section III provides an overview of the genetic privacy ontology; Section IV refines the aspects of the ontology related to consent and restriction enforcement; Section V specifies the rule base using a predicate-based authorization framework, Section VI develops an implementation example with rule definitions and an example focused on obligations, and, finally, Section VII presents conclusions.

II. Related Works

There are existing standards and frameworks with methods to implement various aspects of genetic privacy protections. Integrating the Healthcare Enterprise (IHE) standards profiling organization has developed frameworks, use cases, and specifications for managing the sharing of documents between organizations [19]. The inter-organizational policies must be completed prior to the use of this standard for implementing the consent agreements. There is some but not all the required capability to address components of genetic privacy related to acknowledging consents. For example, the use case of individuals specifying that other specific individuals do, or do not, have access to their data is listed as a scenario that is explicitly not supported [20]. Many state laws call for this type of consent specifications as a prerequisite for permissible access to data. Concepts like Dynamic Consent in biobanks provide opportunities to address the complex requirements inherent in genetic privacy [21]. Dynamic Consent engages the research participant in a real-time personalized process to obtain and update consent as needed. Dynamic Consent is also incorporated into the Bilateral Consent Framework (BCF) which use other techniques and entities such as a trusted mediator to operate the system, auditing, a code of conducts and reputation system to improve and enhance the consent process [22]. However, both Dynamic Consent and BCF have components that would require changes to state laws which often specify how and when consent must be obtained.

The restrictions placed by regulatory environments on information sharing has been identified as an issue that requires coordination across system silos [23]. The Global Alliance for Genomics and Health (GA4GH) provides a framework for sharing genome data with privacy and security policies, technology recommendations, guidance and architecture to allow interactions between organizations [24], [25]. The basis of data sharing in GA4GH is that the donors or their representatives have provided consent in accordance with organizational policies and the applicable laws [26]. The work to date provides comprehensive policies but does not have a functional mechanism for implementing sharing data or addressing the restrictions placed by donors in systems that hold and use such data. The National Institute of Health (NIH) Office of Science Policy (OSP) collaborated with GA4GH to develop a set of consents to improve consistent identification of how genomic data is used [27]. Further work was performed to reclassify these codes into a set of Categories (Primary and Secondary) and Requirements (additional agreements needed for re-use) [28]. The consent code base is focused on research with some additional restrictions than found in State Laws which could be incorporated into the ontology. These codes and NIH processes should be cross-referenced with other standards such as HL7 to provide a more complete representation.

Other health-care privacy ontologies have some overlap with genetic privacy concepts based on laws. However, these ontologies have gaps in numerous areas when compared to implementation requirements of state laws. The HL7 Security and Privacy Ontology has a class PurposeOfUseOntology with a purpose code and description [29]. Because the focus is on health care organizations, the main categories in this ontology are for health care marketing, operations, payment, research, public health and treatment with options for patient requested inquiries including family, power of attorney and support network. This list does not include key purposes regulated by law, such as Law Enforcement, Homeland Security and Insurance access. Other matching HL7 ontologies have some overlap (such as Organization, ObligationPolicy, Refrain, and Role) but not a complete set of genetic information related categories. The Sensitivity class contains a genetic disease information sensitivity but this needs to be set based on the state law attributes of the ontology. Many of the state laws have conditions that must be met prior to releasing genetic information in addition to imposing specific obligations to be adhered to after the release. The Consent component is addressed on a limited scale with options for delegation but not addressing aspects such as capacity. The ontology also has a smaller set of obligations and "refrains" to address some restrictions. A future research option is to develop a mapping and extension between our genetic focused ontology and the HL7 framework as a basis for an implementation.

Genetic privacy protections issues are expanding with the introduction of big data repositories and Direct-to-Consumer (DTC) DNA testing [30]. Adoption of the latter has skyrocketed with its lower prices and wide-spread advertising. DTC DNA testing-related sites encourage sharing of genetic data, including through the use of social media. Naveed et al. provide a Genomic Data Handling Framework to track the protections required from initial collection through to storage and use. The framework groups the uses into the general categories of Healthcare, Research, Legal and Forensics, and DTC along with divisions for the implementation of technical and legal protections. Legal requirements and use cases in

state laws extend the potential areas of use into employment and insurance with legal protections needed throughout the process. In addition, the use lifecycle encompasses the retention and destruction aspects of storage beyond the presented framework.

Another pertinent issue is that consumers often do not have an understanding of the consequences of these DTC services [31]. In general, even when presented with consent agreements, consumers, patients and research participants have a wide variety of reasons for permitting access to their data, do not always fully understand the extent and implications of these agreements, and underestimate the ability for de-identification [7], [32], [33].

Rahmouni et al. developed an ontology of European privacy requirements for sharing patient data between countries [34], [35]. It focused on the implementation of data access between countries with respect to privacy status, consent requirements, recipients, level of detail, purpose, secondary purpose, and access by legal representatives. The consent requirements reflect many of the similar aspects in the US with respect to general areas such as when consent must be obtained, amount of details, written mode, and competency. These are divided up into four classes for necessity, specificity, explicitness, and format. There are no structures for the supplemental requirements prevalent in US laws outside various options for consent agreements and anonymization.

Other healthcare security focused ontologies lack the focus on purpose-driven access found in US laws. Blobel's pHealth has a policy structure that can implement many of the legal requirements and implements patient consent using policies [36]. The patient and internal organizational focus on access policies limits the opportunities to address the wide variety of scenarios prevalent with external access to patient data.

Most privacy models also use Role-Based Access Control (RBAC) to data inquiries and implementing enforcement policies. The use of RBAC has been identified as one of the candidates for implementing privacy access controls in the EHR domain [37], where rights can be assigned based on organizational policies in a hierarchical manner that is modified based on the user's role and then adjusted by the patient as desired. Healthcare privacy extensions, such as those proposed by Hung, provide the structure for adding concepts for areas including purpose, obligations, and retention [38]. The nature of genetic access restrictions and criteria requires a specific framework to accommodate the variations in state laws.

III. GENETIC PRIVACY ENFORCEMENT ONTOLOGY

A. Ontology Overview

The primary components of the genetic privacy ontology based upon Federal and State laws are the Requester, the Request and the Response as seen in Figure 1.

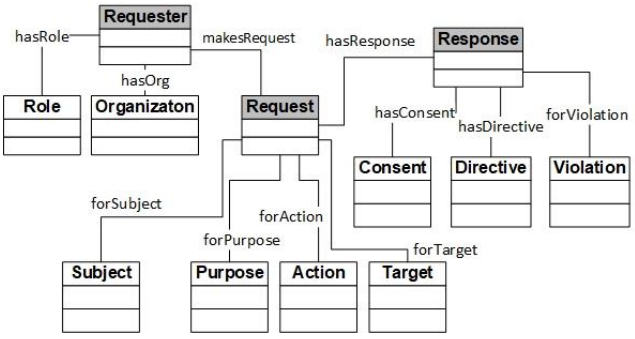


Figure 1. Genetic Privacy Ontology

- **Requester** addresses the person asking for access to the information and associated information such as their role and organization
- **Request** focuses on the purpose the Requester needs the information, the subject of the request (e.g., patient), what specific information is being sought (target), and what action will be performed with the information (e.g., read, retain, update).
- **Response** returns the answer to whether access is permitted or denied along with supplemental requirements for the release including if a consent form is required. The response includes reference material on the potential outcomes regarding a violation where the genetic information is incorrectly handled.

The first two (Requester and Request) are attribute based classes that can be used to generate an access decision (Permit or Deny) based on the Purpose-focused rules. For example, access to genetic information for medical purposes has a different set of permissions and requirements than those for law enforcement. These classes and associated rule base were previously addressed in detail and further information can be found in [18].

Once the access decision is made, then there are potentially a set of other requirements that are not attribute based but still need to be addressed. A Consent Form signed by the subject or their designated representative is often required and is usually generated for each specific request. There are also directives regarding a number of factors included in the state laws to address a large number of areas such as retention, use, supplemental disclosures, and de-identification. These directives may need to be addressed before release as a pre-condition, after release as a restriction on the use of the information, or obligations that the requester must perform after receipt. If the requester fails to adhere to these directives or the consent form directions, then the violation information provides insight into the penalties that can be assessed. The Consent, Directives and Violation classes are the focus of this paper.

B. Ontology Refinement

In our previous works, the ontology has been refined to reflect the increasing level of insights gathered. The first

ontology was developed based on related works review, other existing structures and ontologies, previous research in medical privacy, and the implementation of the laws for several initially selected sample states. The second published version of the ontology reflected adjustments to meet the requirements as more complex state laws were evaluated and compared against the work to date. As described in the methodology section, the efforts then moved to focus entirely on the ontology in order to fully address the state and federal laws specifically related to genetic medical privacy. The third published version reflected these efforts in the areas of Requester and Request classes.

This iteration addresses the criteria and actions to be taken to fulfill the request in accordance with the required conditions and constraints from the applicable Federal and State laws. The previous paper classified these actions into the following groupings:

- **Validations**: Pre-release activities (assuming all other permission criteria have been met) with two subclasses
- **Consen**t: Agreement from the subject or the appropriate representative to release the information along with specific clauses or text that must be included
- **Pre-conditions**: Requirements that must be addressed, completed or agreed to by the information provider or recipient such as ensuring the requester has a need to know
- **Constraints**: Post-release activities that the information recipient must agree to address
- **Restrictions**: Limits of the use, distribution or actions that can be taken with the information such as limiting re-disclosure based on the original purpose
- **Obligations**: Actions the recipient must take after receiving the information such as retention and destruction
- **Penalty**: Potential consequences for violating the validations or constraints associated with a specific rule.

This version of the ontology focusses on completing the analysis and classification of these components. The challenge and a major contribution have been the separation of pre-conditions and obligations. There are numerous instances where requirements are potentially applicable as both a pre-condition and obligation. For example, a "need to know" definitely implies that the current requester must fulfill the requirement as seen in the statement from Illinois Section 30, Disclosure of person tested and test results: Disclosure shall be limited to those who have a need to know the information, and no additional disclosures may be made. However, once the information is provided to the requester, it is a reasonable assumption (and may be required) that the requirement must also be applicable to re-disclosures. Therefore, Obligations and Pre-Conditions have been consolidated into a "Directives" class with attributes for Pre-Condition, Obligation and Restriction to provide additional flexibility during rule development. This change is reflected in the Figure 1.

Consent is a stand-alone class that reflect the specific characteristics of this requirement. While consent is typically thought of as a pre-release requirement, the analysis identified situations where this may occur post-release as well. For example, some states have a requirement that any re-disclosure or use for a different purpose requires additional consents. This condition must be enforced by the initial recipient. As seen in the sections below, notices that must be provided to the subject (or authorized signatory) with the consent are captured as a subclass to Consent. Since there are situations where notices are required based on the results of an information release, such as notices to parents about the results of neo-natal tests, these notices are captured within the Directives class structure.

Finally, some state laws directly state potential penalties associated with inappropriate genetic information release or use in the laws related to this subject. Therefore, a Violation class has been included to provide this information to the requester and reinforce the seriousness of complying with the relevant laws. There are a number of components, such as the violator's intent) to a violation in addition to the assessment of a specific penalty (such as a fine). Therefore, the class label was widened from Penalty to Violation to reflect these other subclasses.

C. Consent Super-Class

The Consent super-class shown in Figure 2 is more complex than simply providing a form for signature. In the analysis process, over 175 statements were extracted from laws regarding consents. State laws dictate a variety of specifications for signatories, format, text, informational notices, supplemental releases, record keeping and when consent is not required. The classes directly associated with the Consent super-class are:

- **Consent Form** represents the actual consent agreement and the associated requirements, directives, classes and notices to be provided. This Consent Form class is decomposed further below to provide information on various aspects.
- **Releases** addresses situations where additional consent may be required prior to the release of information regarding a specific individual or an institution. In the case where the request Purpose was Treatments, attributes flag when a physician must be the consent requester and the need

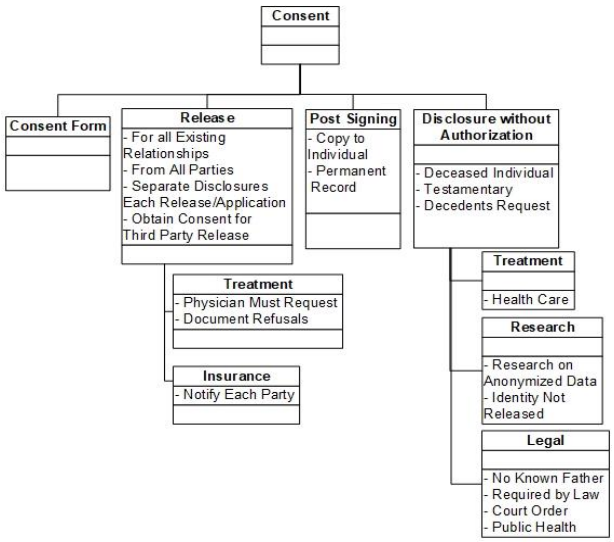


Figure 2. Consent Super-Class.

to document the refusal of a party to sign a consent statement. A requirement for insurance disclosures is to notify all the parties (or their guardian) in a group insurance of the conditions related to requesting genetic information.

- **Post-Signing** directs that the signed consent form must be included in the permanent record and/or a copy provided to the individual.
- **Disclosure Without Authorizations** outlines specific situations where genetic information may be eligible for release without the consent of the individual (or their representative). The three general conditions relate to a deceased individual while the other provide for release for medical treatment, specific types of research and legal conditions.

The Consent Form class in Figure 3 includes a number of attributes that reflect the conditions and requirements for the form itself. Because the focus here is genetic consent agreements, the Consent Form class also has an attribute to reflect the directions from some states that a general information release is not sufficient as seen in Georgia Statue 18.13.010: A general authorization for the release of medical records or medical information may not be construed as the informed and written consent required by this section. In these cases, another consent form that has specific criteria for genetic information must be signed even if there is a general release form on file. The classes associated with Consent Form are as follows:

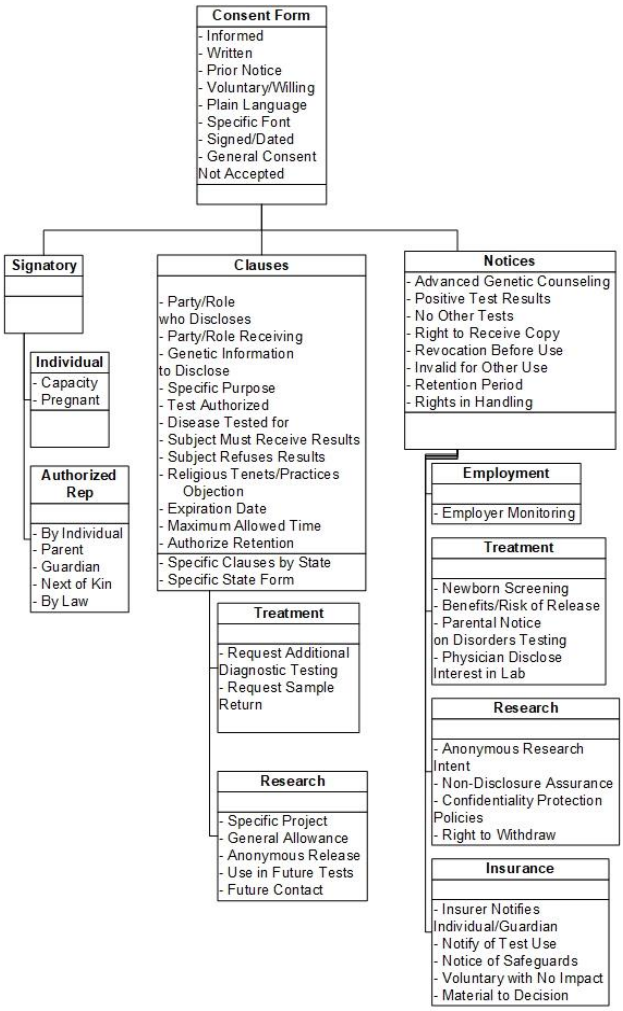


Figure 3. Consent Form Class.

- **Signatory** addresses who can sign the consent form. The individual can sign if they have the capacity (mental or age based). In Texas, a pregnant subject has additional consent requirements for any tests performed on a child in utero (as an implementation detail, these attributes would be reflected in the overall Subject class.) Authorized Representative may sign for individual and this representative may have been designated by the individual to sign on their behalf, may be the parent or guardian, a next of kin if the subject is deceased or set by some other criteria in the law.
- **Requirements** reflects specific statements in the laws regarding the consent form and overall process.
- **Clauses** provides the sections of text that must be included in the consent agreement being signed. Some states require that specific text or forms are used so this option is reflected as additional subclasses.
- **Notices** lists information that must be provided with the consent form. These disclosures provide additional information related to areas such as information use, rights of the subject, potential future use or disposition and participation in specific programs.

D. Directices Super-Class

Once the purpose-based rules are applied for a Requester to gain access based on the Request attributes, the Directives seen in Figure 4 dictate the pre-condition requirements for the release of information (in addition to consents), restrictions that are applied once the information is released, and obligations with specific actions that must be taken after the release. For example, a physician may be allowed to gain access to genetic information based on their role and their participation in a subject's treatment regimen. However, additional requirements might need to be addressed such as whether the physician has a need to know genetic information to provide care in the current use case. An Emergency Room doctor trying to access genetic information while treating a laceration might have to assert their Need to Know prior to gaining access. (Inappropriately asserting the need then becomes an external audit function and subject to the potentially involves the Penalty class.) These directives may not be retro-active so the date genetic information was obtained may be applicable in some states and use cases. These restrictions have been grouped into the following classes:

- **Disclosure** generally constrains the use of genetic information once permission rules are validated. (Releasing sub-sets of genetic information is addressed in the Limited class.) While some states may prevent additional disclosure under this request, other states provide specific criteria where re-disclosure is permitted which are addressed in a subclass.
- **Limited** provides criteria where disclosure of specific genetic information may be appropriate. AS opposed to Disclosure class, the Limited class addresses the ability for

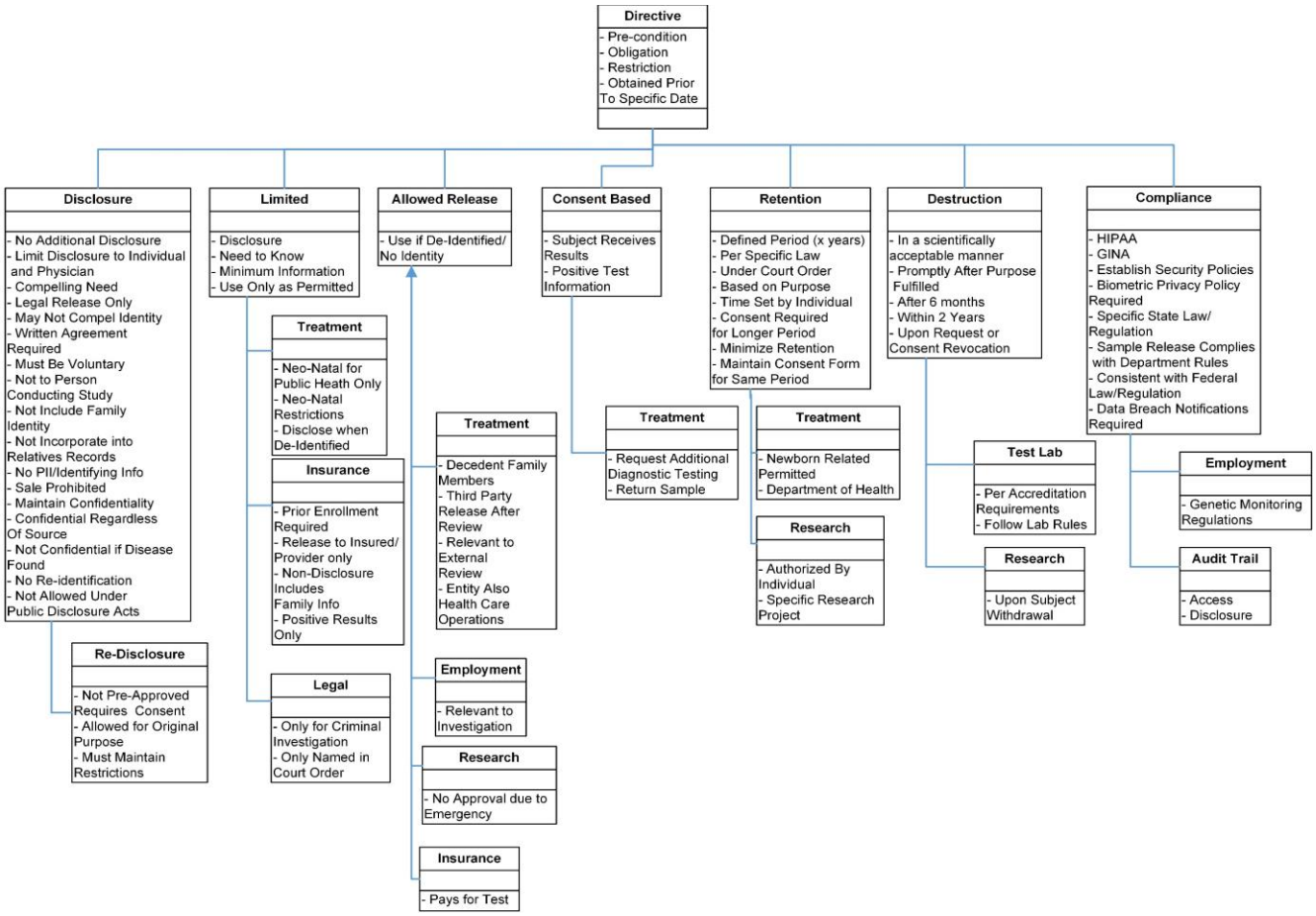


Figure 4. Directive Super-Class.

Disclosure to be for only a portion on the overall genetic information. Specific subclasses deal with situations that are only relevant to Treatment, Insurance or Legal based on the wording in the associated laws.

- **Allowed Release** addresses use cases where the law specifically states that information can be released. If Consent is required for an Allowed Release use case, the combination of Consent and Allowed Releases instances would be articulated in the relationships with the Release super-class.
- **Consent Based** reflects obligations and directives that must be enforced based on clauses and directives in the consent form.
- **Retention** specifies how long the genetic information may be retained either as specific time periods or general guidance. As an example of relationships between classes, Retention can be set by an individual with Authorize Retention as an option under Consent class.
- **Destruction** provides further direction on what to do when the retention period expires or upon a specific trigger. Test Labs have additional directions to follow practices determined by their accreditation requirements or lab guidelines. If a subject withdraws from a research project, destruction may be required depending on the state
- **Compliance** articulates state guidance to comply with certain laws, regulations or rules along with requiring policies and rules to be set related to the release of genetic information. In order to demonstrate compliance, some states provide specific guidance on recording access and disclosing audit records.

E. Violation Super-Class

As stated above, the Violation super-class shown in Figure 5 involves several aspects of addressing inappropriate release of genetic information as specific by Federal and State Laws. The following classes are under this category:

- **Intent** provides information on why the violator released the information. While the release may be Inadvertent, the penalties may be reduced as compared to a willful release.
- **Basis** describes the reasons or methods for the violation. These range from some type of gain (personal, corporate or otherwise) to wanting to inflect harm on someone.
- **Offense** separates out the criminal from non-criminal violations.
- **Actions** indicates how the violation will be addressed primarily for non-criminal offenses. (Criminal

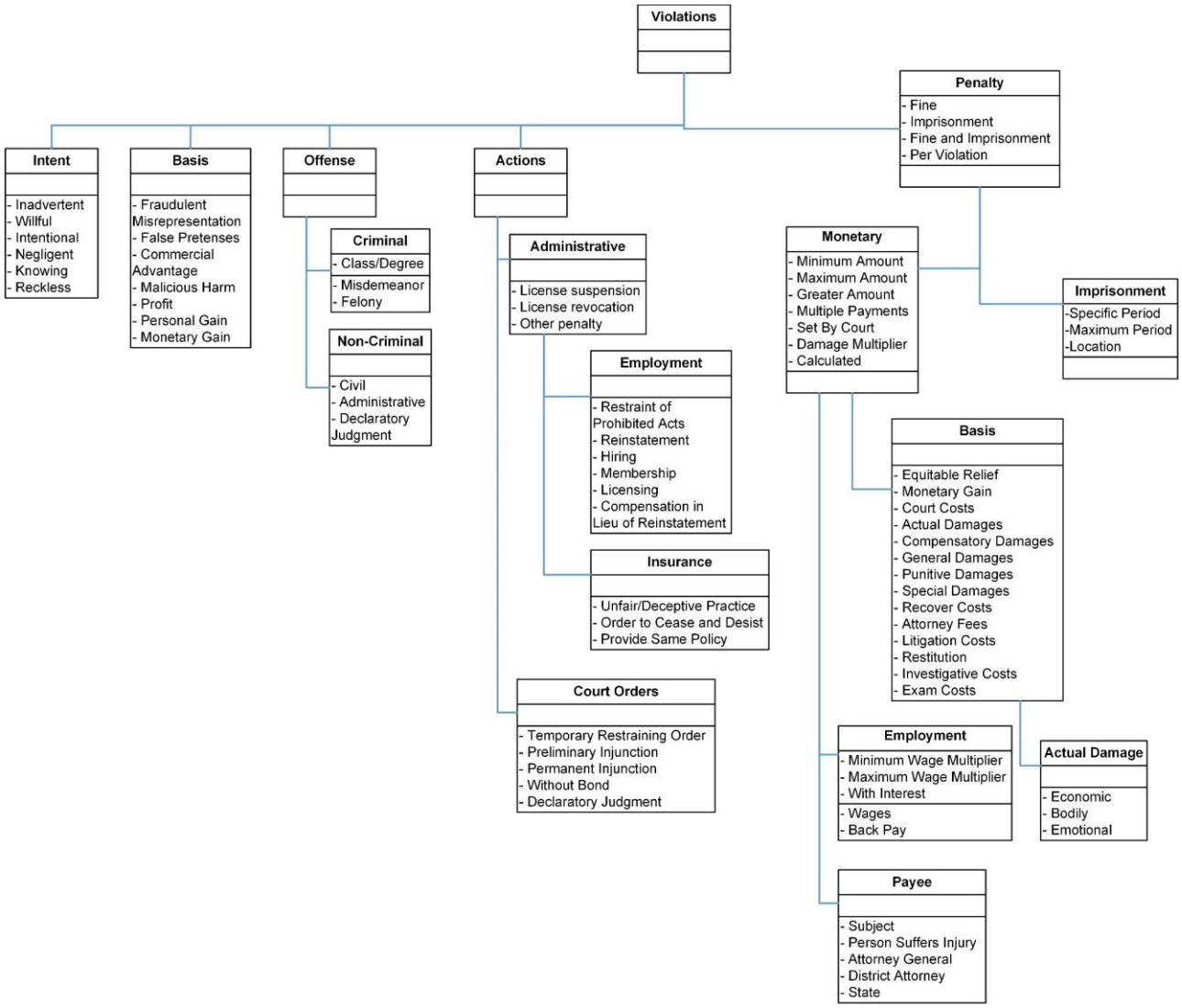


Figure 5.Violation Super-Class.

offenses may also trigger these actions but are not the primary focus in state laws.)

- **Penalty** indicates what the potential outcomes can be if a criminal offense is confirmed. The Penalty may be Monetary, Imprisonment, or a combination and there may be a penalty assessed for each violation.
  - **Monetary** has a number of attributes that may be associated with the assessed amount. In some cases, the amount may be set depending on a specific Basis. There are some penalties associated with Employment and the use of wages as a computational component. The Payee may be the Subject or some other party if not reimbursed back to the Government.
  - **Imprisonment** is usually set as a specific period to be served (in months or years) or a maximum amount. The location may be designated as a County or State facility.

F. Logical Definition Formulation

We provide a logical formulation to articulate the complexities of the genetic privacy protections in consideration of the wide variety of attributes and formulations that need to be specified in a hierarchical manner [39].

The core of the framework is a 5-tuple Data System (OTH, UGH, RH, A, Rel) where the elements are:

- OTH is an object-type hierarchy which in this case is an Electronic Health Record
- UGH is a user group hierarchy representing the system membership in the EHR
- RH is a role hierarchy for the role-based access permissions prevalent in the genetic information access requirements. For example, the HealthCare Provider role has subordinate roles such as the Physician and Nurse roles.

- A is the authorization nodes or Actions that can be performed on objects and is represented in the ontology as the Action super-class
- Rel is the set of relationships that links together element using unary, binary or n-ary tuples. For example, a medical record created by a physician is represented using a DidCreate(record, user) and thus will be provided additional access rights under the authorization component.

Within FAF, an authorization rule is in the form

$$\text{head}(o, s, \langle sign \rangle a) \leftarrow L_1 \& \cdots \& L_n.$$

where o is an object, s is a subject performing the actions, a is an action with the sign (+ or -) indicating permission or denial, and $L_1, \ldots, L_n$ are done, hie-, or rel-literals. done is a predicate stored in the history table, hie-literals are hierarchy predicates and rel-literals are application specific predicates. The rel-predicates provide the vehicle to address the majority of specifications found in the rule base.

However, within the context of accessing genetic information in an EHR, the cando tuple needs to be expanded to address the complexities found in a real-world situation. Therefore, cando is updated into shareable() to include the individual (i) patient's capacity to consent, the request (r) being made to obtain the purpose and state for the physical location, and the consent form (c) to determine if the individual has provided access. The revised formulation is

$$\text{shareable}(o,s,i,r, \langle sign \rangle a) \leftarrow L_1 \& \cdots \& L_n.$$

The set of rel-predicates includes:

- AllowedRole(r, s)
- AllowedOrg (r, s)
- PurposeAccess (o, a, r)
- needConsent(r,i)
- haveConsent (r,i)
- hasDirective (r, s)
- hasObligation (r, s)
- Precondition (r, s)
- WrittenForm (r)

If NeedConsent evaluates to true, then the haveConsent predicate is evaluated based on the following:

haveConsent() ← giveConsent(r,i), ConsentOnFile(r,i)
noConsent() ← ¬giveConsent(r,i), ¬ConsentOnFile(r,i)

The final decision is evaluated as follows to address any conflicts:

finalConsentDecision() ← haveConsent(), ¬ noConsent()

In addition, the negation of a predicate indicates that some aspect is not needed within this context. For example, if consent is not required (as is common for law enforcement requests), then ¬NeedConsent(r,i) indicates that the individual's consent is not required for this request.

In a similar manner to cando in FAF, dercando is updated to dershareable to implement two areas where authorization is derived from inferences. giveConsent can potentially be a hierarchical situation where multiple consents may be present that must be evaluated.

For example, in some cases a minor child can give permission without their parent's approval. Alternatively, there are other scenarios where the parents can provide consent that overrides the minor's directives. In addition, permissions granted to one part of a medical record will be inherited by related or "lower" aspects of the record. The person who requested a genetic test is often provided access to the test results by inheritance. If consent is provided to all genetic information within the record to the physician, it is obvious this permission is inherited by all genetic-related information.

IV. IMPLEMENTATION

Our previous work included a prototype that uses a workflow engine to gather the required attributes, display the results and confirm the implementation of the directives [17], [18]. These papers include additional information on the specific steps in the workflow and the overall development process along with screen shots and the rules algorithm.

The workflow engine, Yet Another Workflow Language (YAWL), is compatible with Electronic Health Record (EMR) systems such as OpenMRS [40]. This integration will allow many of the attributes about the request and requester to be directly extracted from the record repository. In addition, the prototype will force a consent to be obtained in case of need. As seen in Figure 6, these attributes are then extracted by the Consent Service and the Protégé ontology populated for the execution of the rules. The results are processed by the Rules Hierarchy Algorithm, and Consent Service to develop the Final Access Results. The results also include any associated pre-conditions, obligations, restrictions and violations that are part of the enforcement component. These results are returned to the workflow for display of the access decision along with a confirmation that the consent has been completed along with agreement the directives have been evaluated and addressed.

The workflow process is shown in Figure 7. Section 1 of the workflow gathers the required attributes and generates the

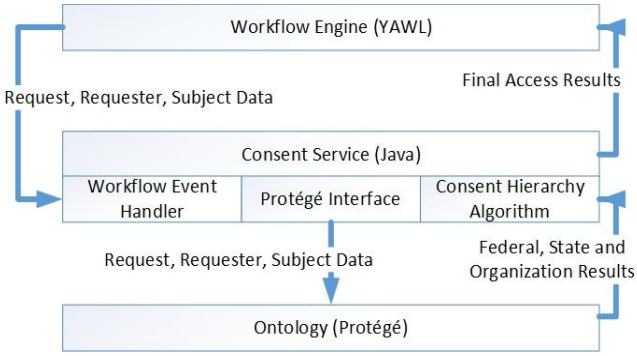


Figure 6. Prototype Architecture.

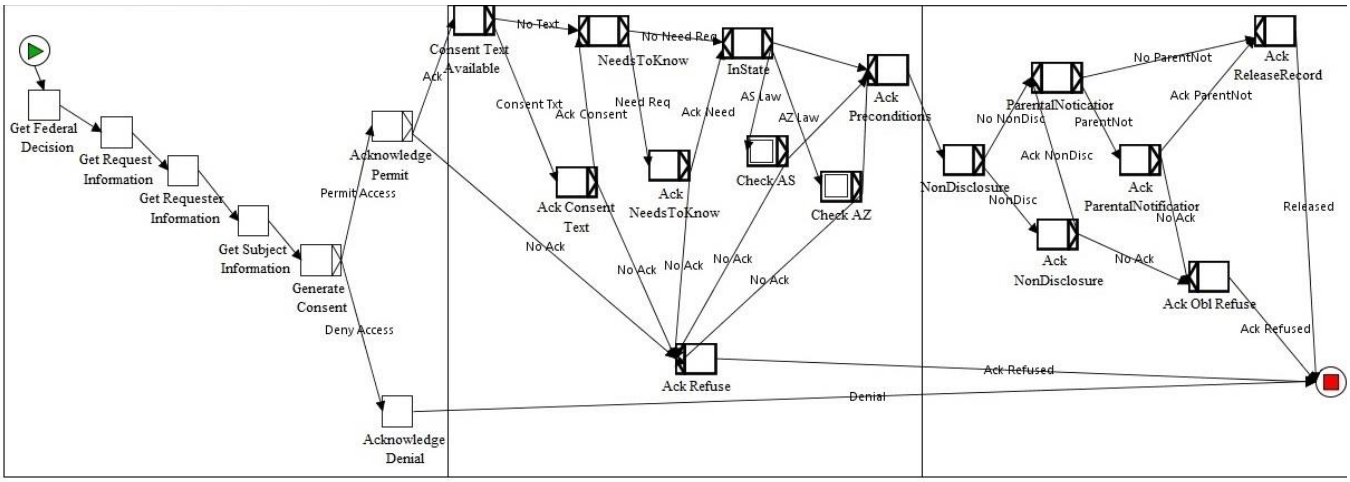


Figure 7. Prototype Workflow.

access decision. Section 2 displays the complete set of results and requests a confirmation. The next set of steps generates the consent agreement and requests an electronic signature. Once the consent agreement is signed, the directives identified as pre-conditions are displayed and the requester validates that the conditions have been met. Once the pre-conditions are met, the restrictions and obligations are provided to be enforced as part of the release process. Failing to complete any of the steps or provide the necessary signatures/agreements in Sections 2 and 3 will result in the access decision converting to a Deny.

A. SWRL Rule Structure

This section specifies the rule structure for all access inquiry rules and the associated release conditions in the Protégé prototype [16], [17]. The workflow captures the required data elements and the Consent Service populates the Protégé ontology instances with the submitted data. The workflow has three steps to collect the access inquiry information on the request, requester and subject as described in our previous papers. The SWRL rule structure defines how the ontology-based rules are built in order to provide an access decision along with any applicable response details for consent, directives and violations. (The structure is provided in Figure 1.)

Basic Structure

At a minimum, each rule must have a request, requester and response. In addition, any required response condition instances (consent, directives and violations) are retrieved in the precedent for use in the antecedent.

- **g(inquiry), f(request), f(requester), g(response), g(conditions) -> s(response), s(conditions)**

The g() predicates get the instances from the ontology for evaluation or use in the response. The f() predicates then evaluate the instances to determine if the rule is applicable for this access request. The s() functions in the antecedent sets the response instance attributes and associates the retrieved conditions with the response.

g(inquiry) retrieves the required instances present in all rules for both the request and requester instances with one statement.

- **makesRequest(?r, ?req**) which provides the linked request and requester instances.

f(request) and f(requester) provide the predicates for evaluating if the rule is applicable to the access inquiry. Since these two functions encompass the entire rule base, selected functions are provided to illustrate their operation.

One example is to restrict the rule enforcement to a specific state. The following predicate operates on an attribute in the request instance.

- **inState(?req, abbr)** where abbr is the two letter state abbreviation such as “DE” for Delaware.

If the rule is enforcing a constraint in a subclass to the requester (role or organization) or request (subject, purpose, action or target), the function first retrieves the associated instance and then performs the evaluation.

- **forResource(?req,?resource), isGeneticResult(?resource, true)** which is used to determine the specific part of the medical record that is being accessed (forResource) and if the record component contains genetic information based on a resource attribute (isGeneticResult).
- **forPurpose(?req, ?pur), isTreatment (?pur, true)** where an Purpose instance (forPurpose) attribute (isTreatment) is evaluated to determine if the purpose is in a specific grouping. In this example, all medically oriented treatment purposes are grouped in the isTreatment property on the Purpose.

g(response) is comprised of two SWRL statements because there are three separate response objects that can be associated with each access inquiry for the Federal, State and Organization levels. The objects are respectively ?res, ?resst and ?resorg to match the three levels. Each rule only contains one response instance and the correct response level is obtained using one of these combinations:

- **hasResponse(?req, ?res), responseLevel(?res, "Federal")**

- **hasResponse(?req, ?resst), responseLevel(?resst, "State")**
- **hasResponse(?req, ?resorg), responseLevel(?resorg, "Org")**

g(condition) gets the instances that will be used in the antecedent to establish the conditions associated with the information release. The instances are obtained from the Consent, Directive and Violation super-classes and all subclasses that reflect the release requirements. For example, one requirement often imposed by states is that the requester must have a "need to know" in order to be permitted access to the genetic information. In the first step, the "Need to know" instance is retrieved in the following rule snippet. Under s(condition) the instance is associated with the state response.

- **oblName(?pre, "NeedToKnow")** is used to retrieve the Directive instance with the name "NeedToKnow" and then associate this instance with ?pre. ?pre is used in s(response) to enforce the requirement.

Multiple conditions are associated with one rule by creating unique instances to replace ?pre in the formula. For example, ten consent clauses can be associated with one rule by replacing ?pre with ?clause1…?clause10.

s(response) sets the response instance attributes to reflect if access is permitted and provide supporting information on the basis of the access decision. The assignment statements are:

- **isAllowed(?resst, boolean)** to set the access decision to permit (true) or deny (false)
- **canOverride(?resst, boolean)** to communicate if "lower" level rules can override this rule. For example, if the State law permits access, an override of false means the organization can't deny access.
- **decisionSource(?resst, text)** to provide the reference information for the rule from the applicable law, regulation or policy.
- **hasRule(?resst, integer)** is an implementation specific construct to simplify debugging and gives each rule a unique ID

s(condition) uses relationship statements to associate the previously retrieved conditions with the response instance.

- **hasPreCondition(?resst, ?pre)** takes the ?pre instance retrieved for the NeedToKnow example and uses hasPreCondition to associate the condition with the ?resst release instance obtained under g(release).

Simple Rule Example

In this SWRL rule, a person is allowed access to their own medical records for information regarding genetic test results. This rule implements a portion of a Georgia State Law: Information derived from genetic testing shall be confidential and privileged and may be released only to the individual tested.

makesRequest(?r, ?req), isSelf(?r, true), inState(?req, ""GA""), forResource(?req, ?resource), isGeneticResult(?resource, true), forPurpose(?req, ?pur), purposeDesc(?pur, ""SelfRequest""), hasResponse(?req, ?resst), responseLevel(?resst, ""State"") -> isAllowed(?resst, true), canOverride(?resst, false), decisionSource(?resst, ""GA LAW 33-54-3""), hasRule(?resst, 10)"

g(inquiry): makesRequest(?r, ?req)
f(request): inState(?req, ""GA""),
forResource(?req, ?resource),
isGeneticResult(?resource, true),
forPurpose(?req, ?pur), purposeDesc(?pur, ""SelfRequest""),
f(requester): isSelf(?r, true),
g(response): hasResponse(?req, ?resst),
responseLevel(?resst, ""State""),
s(response): isAllowed(?resst, true),
canOverride(?resst, false),
decisionSource(?resst, ""GA LAW 33-54-3""),
hasRule(?resst, 10)"

Note there are no conditions associated with this rule since the law does not impose any restrictions.

V. RESPONSE EXAMPLE

The following snippets from Delaware State Law Chapter 12, Informed Consent and Confidentiality, provides an example that would be applicable to a request for medical treatment:

> *(a)No person shall obtain genetic information about an individual without first obtaining informed consent from the individual.*
>
> *(4) Informed consent"*
>
> *a. For the purpose of obtaining genetic information, means the signing of a consent form which includes a description of the genetic test or tests to be performed, its purpose or purposes, potential uses, and limitations and the meaning of its results, and that the individual will receive the results unless the individual directs otherwise;*
>
> *(a) Regardless of the manner of receipt or the source of genetic information, including information received from an individual, a person shall not disclose or be compelled, by subpoena or any other means, to disclose the identity of an individual upon whom a genetic test has been performed or to disclose genetic information about the individual in a manner that permits identification of the individual, unless…:*
>
> *(b) Any person who willfully obtains or discloses genetic information in violation of this subchapter shall be punished by a fine not less than $5,000 not more than $50,000.*

A. SWRL Rule

A sample SWRL rule that would be invoked for this scenario is as follows:

makesRequest(?r, ?req), inState(?req, "DE"), forResource(?req, ?resource), isGenetic(?resource, true), forPurpose(?req, ?pur), isTherapeutic(?pur, true), hasResponse(?req, ?resst), responseLevel(?resst, "State"), oblName(?consent, "ConsentRequired"), oblName(?dir, "MayNotCompelIdentity"), vioName(?vio, "Willfull") -> isAllowed(?resst, true), canOverride(?resst, false), hasConsent(?resst, ?consent), isSigned(?consent, true), isDescription (?consent, true), isPurpose(?consent, true), ?isUse(?consent, true),

?isPositiveTestResults(?consent, true), isSubjectReceivesResults(?consent, true), hasDirective (?resst, ?dir), forViolation(?resst, ?vio), isMin(?vio, 5000), isMax(?vio, 50000), decisionSource(?resst, "DE LAW 12"), hasRule(?resst, 1200)

In this rule,

- **?r** is for the Requester of the Request
- **?req** is for the Request that links the various components, such as Subject, Purpose and Resource
- **?pur** is the Therapeutic Purpose that is associated with the Request
- **?resource** is for the "GeneticTestResults" part of the medical record
- **?consent** is for the Consent Required clause
- **?dir** is the Directive that the recipient may not be compelled to reveal the subject's identity.
- **?vio** is the Willful Violation object
- **?resst** is the State Response object that is associated with the Request.

These SWRL statements are explained in Table I.

TABLE I. SAMPLE RESPONSE STATEMENT RULE

SWRL Statement	***Explanation***
makesRequest(?r, ?req)	Links Requester for the Request
inState(?req, "DE")	Request is for Delaware
forResource(?req, ?resource)	Links Request with the Resource
isGenetic(?resource, true	Restricts the rule to a Resource that is identified as a genetic information
forPurpose(?req, ?pur)	Links Request with Purpose
isTherapeutic(?pur, true)	Restricts the rule to a Treatment Purpose
hasResponse(?req, ?resst)	Links the Request with a Response to store answer
responseLevel(?resst, "State")	Gets the Response for State level to store answers
oblName(?consent, "ConsentRequired")	Gets the Consent Required Object
oblName(?dir, "MayNotCompelIdentity")	Gets the appropriate Directive object
vioName(?vio, "Willfull")	Gets the appropriate Violation object
-> isAllowed(?resst, true)	Sets the State response to access is allowed
canOverride(?resst, false)	Sets the state Response to not allow override by organization
hasConsent(?resst, ?consent)	Sets the State response to include the Consent Required condition
isSigned(?consent, true), isDescription (?consent, true), isPurpose(?consent, true), ?isUse(?consent, true), ?isPositiveTestResults(?consent, true), isSubjectReceivesResults(?consent, true)	Sets the attributes on the Consent object to require a signed form that includes the description, purpose, use, test results meaning and subject receiving results
hasDirective (?resst, ?dir)	Sets the State response to include the MayNotCompelIdentity obligation
forViolation(?resst, ?vio)	Sets the State response to include the Wilful Violation
isMin(?vio, 5000), isMax(?vio, 50000)	Sets the Minimum and Maximum Fine amounts for the Violation
decisionSource(?resst, "DE LAW 12")	Sets the State response to reflect the decision source as state law
hasRule(?resst, 1200)	Sets the rule number to 1200 for reference

B. Obligation Enforcement

Once the genetic information has been released, obligations then require additional interactions in order to enforce the applicable laws, regulations and consent directives. A YAWL workflow is presented in Figure 8 to address the required process. A trigger in the underlying EHR invokes the workflow and creates a workflow item for evaluation. A workflow path then is selected from the following list to execute the associated rules:

- **Re-disclosure Request** determines if further dissemination is permitted based on the original request. There are three identified conditions where additional consent may be required, the re-disclosure can be performed if the purpose is the same as the original request, and only if the original restrictions are agreed upon for the release.
- **Test Results Received** from a genetic test request is evaluated for specific conditions associated with the original request and consent form. Subject notification may be required that the results were received. As a separate workflow, the subject may be entitled to additional information if the test is positive for the condition.
- **Retention Expired** begins the review process to determine if the genetic information/sample can be retained longer. If the criteria are met, then the retention period can be extended. Otherwise the information and/or sample is subject to destruction. For samples, the destruction may be subject to

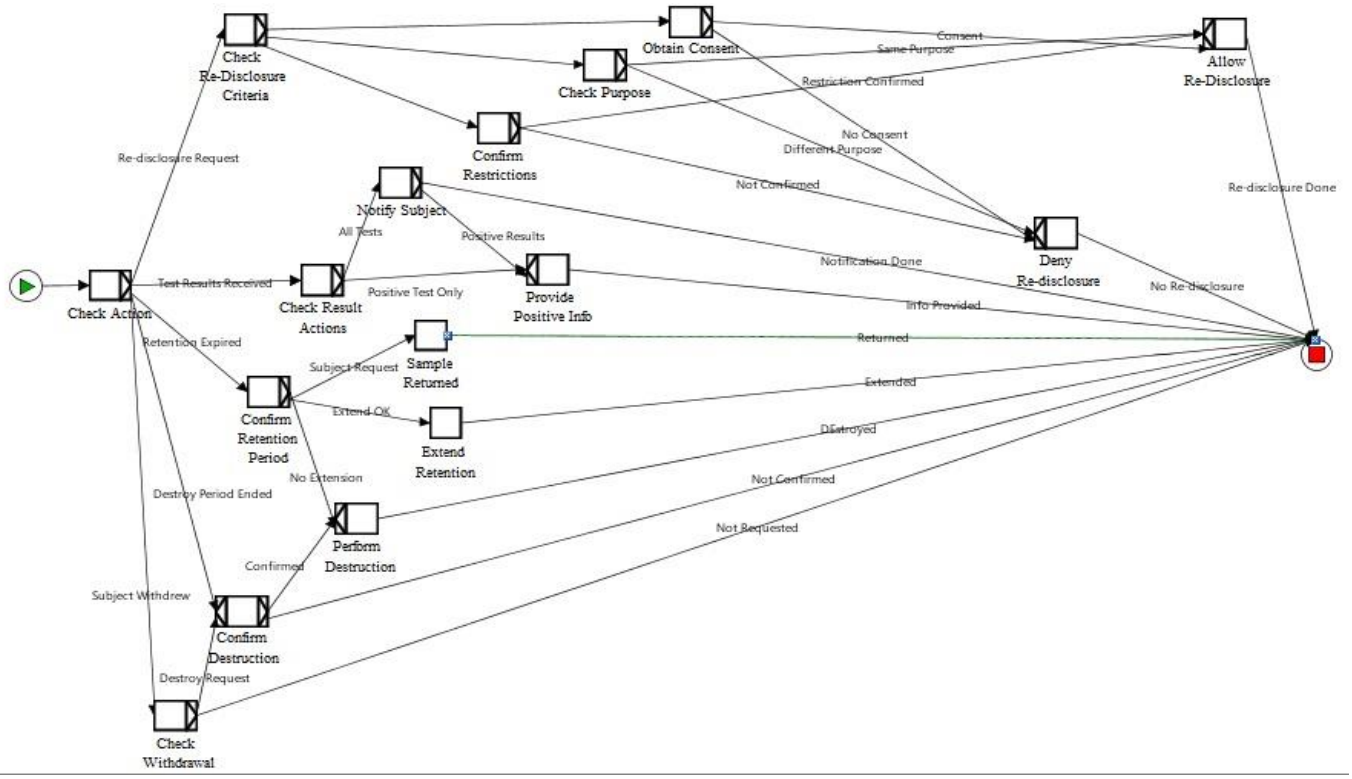


Figure 8. Obligation Workflow.

specific requirements or, if stated in the consent agreement stated, the sample must be returned to the subject.

- **Destroy Period Ended** indicates the time has expired and the sample must be destroyed in accordance with any specified processes.
- **Subject Withdrew** from a research project may impact the sample and trigger a destruction action.

C. Obligation Example

The following snippets from Nevada State Law Chapter 629.161, Retention of genetic information that identifies person without consent unlawful; exceptions; destruction of genetic information, provides an example that would be applicable to destruction for a subject withdrawing from a research study:

> *3. Except as otherwise provided in subsection 4 or by federal law or regulation, a person who obtains the genetic information of a person for use in a study shall destroy that information upon:*
>
> *(a) The completion of the study; or*
>
> *(b) The withdrawal of the person from the study, whichever occurs first.*
>
> *4. A person whose genetic information is used in a study may authorize the person who conducts the study to retain that genetic information after the study is completed or upon his or her withdrawal from the study.*

The following snippets from Nevada State Law Chapter 629.191, Penalty, indicates the outcome for a failure to comply with the previous clause:

> *A person who violates any of the provisions of NRS 629.151, 629.161 or 629.171 is guilty of a misdemeanor.*

A sample SWRL rule that would be invoked for a subject that has not agreed to have the information retained upon withdrawal from a research study is as follows:

madeRequest(?r, ?req), inState(?req, "NV"), forResource(?req, ?resource), isGenetic(?resource, true), forPurpose(?req, ?pur), isResearch(?pur, true), isResearch(?resource, true), forSubject(?sub, ?req), hasAction(?sub, "WithDrew"), hasConsent(?sub, ?consent), hasRetainWithdraw (?consent, ?clause), isRetainAllowed(?clause, false), responseLevel(?resst, "State"), oblName(?obl, "DestroyInfo"), degreeName(?degree, "Misdemeanor") -> hasObligation (?resst, ?obl), forViolation(?resst, ?degree) , decisionSource(?resst, "NV LAW 629.161"), hasRule(?resst, 1201)

In this rule, the following instances are added from the previous example:

- ?sub is for the Subject of the Request
- ?consent represents the Consent Agreement with the Subject
- ?clause are the clauses in the Consent Agreement

• ?obl is the obligation that the organization must fulfill

• ?degree provides the class/degree associated with this criminal offense if the information is not destroyed

These SWRL statements are explained in Table II.

TABLE II. SAMPLE OBLIGATION RULE

SWRL Statement	***Explanation***
madeRequest(?r, ?req)	Links Requester for the original Request
inState(?req, "NV")	Request is for Nevada
forResource(?req, ?resource)	Links Request with the Resource
isGenetic(?resource, true	Restricts the rule to a Resource that is identified as a genetic information
forPurpose(?req, ?pur)	Links Request with Purpose
isResearch(?pur, true)	Restricts the rule to the Research Purpose
forSubject(?sub, ?req)	Obtains the Subject associated with the Request
hasAction(?sub, "WithDrew")	Indicates the Subject has the Action for Withdrew
hasConsent(?sub, ?consent)	Obtains the Consent Agreement for this Subject
hasRetainWithdraw (?consent, ?clause)	Restricts the Rule to the Subject having the Clause to Retain Information upon Withdrawal
isRetainAllowed(?clause, false)	Determines that the Retention is set to false
hasResponse(?req, ?resst)	Links the Request with a Response to store answer
responseLevel(?resst, "State")	Gets the Response for State level to store answers
oblName(?obl, "DestroyInfo")	Gets the Obligation to Destroy Info
degreeName(?degree, "Misdemeanor")	Gets the appropriate degree object
->hasObligation(?resst, ?consent)	Sets the State response to Destroy Info
forViolation(?resst, ?degree)	Sets the State response as the Misdemeanour Violation
decisionSource(?resst, "NV LAW 629.161")	Sets the State response to reflect the decision source as state law
hasRule(?resst,1201)	Sets the rule number to 1201 for reference

Upon receiving this result, the workflow would advance to the destruction path to implement the clauses in the subject's consent agreement.

VI. CONCLUSION AND FUTURE WORK

Our genetic privacy ontology was built directly from the applicable Federal and State laws without any pre-conceived boundaries or required elements. The work demonstrates the importance of a purpose-focused structure to appropriately link the various data elements necessary to permit or deny access to the genetic medical information. The ontology and previous prototype work allows the data collection to be directly integrated into EHRs. The next step will be validating an integrated EHR, ontology and prototype using operational data and genetic data requests to demonstrate the appropriate data protections are enforced. This comprehensive integration reduces the provider's effort and provides access decisions in accordance with relevant laws, policies and regulations.

REFERENCES

[1] M. Reep, B. Yu, D. Wijesekera, and P. Costa, "An Ontology for Specifying Regulation-Compliant Genetic Privacy Policies," in *eTELEMED 2018, The Tenth International Conference on eHealth, Telemedicine, and Social Medicine*, 2018, pp. 66–74.

[2] K. Wuyts, R. Scandariato, G. Verhenneman, and W. Joosen, "Integrating Patient Consent in e-Health Access Control," *Int. J. Secure Softw. Eng.*, vol. 2, no. 2, pp. 1–24, 32 2011.

[3] J. Pritts, "The Importance and Value of Protecting the Privacy of Health Information: The Roles of the HIPAA Privacy Rule and the Common Rule in Health Research." Institute of Medicine, 2008.

[4] M. H. Ullman-Cullere and J. P. Mathew, "Emerging landscape of genomics in the electronic health record for personalized medicine," *Hum. Mutat.*, vol. 32, no. 5, pp. 512–516, May 2011.

[5] M. D. Ritchie, E. R. Holzinger, R. Li, S. A. Pendergrass, and D. Kim, "Methods of integrating data to uncover genotype-phenotype interactions," *Nat. Rev. Genet.*, vol. 16, no. 2, pp. 85–97, Feb. 2015.

[6] C. Pihoker *et al.*, "Prevalence, Characteristics and Clinical Diagnosis of Maturity Onset Diabetes of the Young Due to Mutations in HNF1A, HNF4A, and Glucokinase: Results from the SEARCH for Diabetes in Youth," *J. Clin. Endocrinol. Metab.*, vol. 98, no. 10, pp. 4055–4062, Oct. 2013.

[7] Y. M. Lee, R. P. McKillip, B. A. Borden, C. E. Klammer, M. J. Ratain, and P. H. O'Donnell, "Assessment of patient perceptions of genomic testing to inform pharmacogenomic implementation," *Pharmacogenet. Genomics*, vol. 27, no. 5, pp. 179–189, May 2017.

[8] B. M. Knoppers, "Genetic information and the family: are we our brother's keeper?," *Trends Biotechnol.*, vol. 20, no. 2, pp. 85–86, Feb. 2002.

[9] S. Dheensa, A. Fenwick, S. Shkedi-Rafid, G. Crawford, and A. Lucassen, "Health-care professionals' responsibility to patients' relatives in genetic medicine: a systematic review and synthesis of empirical research," *Genet. Med.*, vol. 18, no. 4, pp. 290–301, Apr. 2016.

[10] S. B. Trinidad, T. B. Coffin, S. M. Fullerton, J. Ralston, G. P. Jarvik, and E. B. Larson, "'Getting off the Bus Closer to Your Destination': Patients' Views about Pharmacogenetic Testing," *Perm. J.*, vol. 19, no. 3, pp. 21–27, 2015.

[11] W. W. Lowrance, "Privacy, Confidentiality, and Identifiability in Genomic Research." Discussion document for workshop convened by the National Human Genome Research Institute, Bethesda, Oct-2006.

[12] M. Gymrek, A. L. McGuire, D. Golan, E. Halperin, and Y. Erlich, "Identifying Personal Genomes by Surname Inference," *Science*, vol. 339, no. 6117, pp. 321–324, Jan. 2013.

[13] O. for C. Rights (OCR), "Privacy Rule General Overview," *HHS.gov*, 05-Nov-2015. [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/general-overview/index.html. [Accessed: 01-Jun-2019].

[14] O. for C. Rights (OCR), "Genetic Information," *HHS.gov*, 07-May-2008. [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/special-topics/genetic-information/index.html. [Accessed: 01-Jun-2019].

[15] A. A. Lemke *et al.*, "Patient perspectives following pharmacogenomics results disclosure in an integrated health system," *Pharmacogenomics*, Feb. 2018.

[16] "State and Federal Consent Laws Affecting Interstate Health Information Exchange." [Online]. Available: https://classic.nga.org/cms/home/nga-center-for-best-practices/center-divisions/center-issues/page-health-issues/data-and-analytics/col2-content/list--right/resources/content-reference@/state-and-federal-consent-laws-a.default.html. [Accessed: 01-Jun-2019].

[17] M. Reep, B. Yu, D. Wijesekera, and P. Costa, "Sharing Data under Genetic Privacy Laws," in *Proceedings of the Eleventh Conference on Semantic Technology for Intelligence, Defense, and Security*, Fairfax VA, USA, 2016, pp. 46–54.

[18] M. Reep, B. Yu, D. Wijesekera, and P. Costa, "Sharing Genetic Data under US Privacy Laws," in *11th International Joint Conference on Biomedical Engineering Systems and Technologies*, Funchal, Madeira, Portugal, 2018, vol. 5: HEALTHINF, pp. 349–360.

[19] "Enabling Document Sharing Using IHE Profiles - IHE Wiki." [Online]. Available: http://wiki.ihe.net/index.php/Enabling_Document_Sharing_Using_IHE_Profiles. [Accessed: 01-Jun-2019].

[20] "Basic Patient Privacy Consents - IHE Wiki." [Online]. Available: http://wiki.ihe.net/index.php?title=Basic_Patient_Privacy_Consents. [Accessed: 01-Jun-2019].

[21] J. Kaye, E. A. Whitley, D. Lund, M. Morrison, H. Teare, and K. Melham, "Dynamic consent: a patient interface for twenty-first century research networks," *Eur. J. Hum. Genet.*, vol. 23, no. 2, pp. 141–146, Feb. 2015.

[22] Y. Erlich *et al.*, "Redefining Genomic Privacy: Trust and Empowerment," *PLOS Biol.*, vol. 12, no. 11, p. e1001983, Nov. 2014.

[23] T. G. A. for G. and Health*, "A federated ecosystem for sharing genomic, clinical data," *Science*, vol. 352, no. 6291, pp. 1278–1280, Jun. 2016.

[24] "Global Alliance for Genomics and Health: Privacy and Security Policy." Global Alliance for Genomics and Health, 26-May-2015.

[25] "Standards and implementation practices for protecting the privacy and security of shared genomic and clinical data." Global Alliance for Genomics and Health, 09-Aug-2016.

[26] B. M. Knoppers, "Framework for responsible sharing of genomic and health-related data," *HUGO J.*, vol. 8, no. 1, p. 3, Dec. 2014.

[27] admin, "Protecting Data, Promoting Access: Improving Our Toolbox," *Office of Science Policy*, 02-May-2016.

[28] S. O. M. Dyke *et al.*, "Consent Codes: Upholding Standard Data Use Conditions," *PLoS Genet.*, vol. 12, no. 1, Jan. 2016.

[29] "HL7 Version 3 Standard: Security and Privacy Ontology, Release 1." Health Level Seven International, May-2014.

[30] M. Naveed *et al.*, "Privacy in the Genomic Era," *ACM Comput Surv*, vol. 48, no. 1, pp. 6:1–6:44, Aug. 2015.

[31] H. Shen and J. Ma, "Privacy Challenges of Genomic Big Data," in *Healthcare and Big Data Management*, Springer, Singapore, 2017, pp. 139–148.

[32] J. M. Oliver, M. J. Slashinski, T. Wang, P. A. Kelly, S. G. Hilsenbeck, and A. L. McGuire, "Balancing the Risks and Benefits of Genomic Data Sharing: Genome Research Participants' Perspectives," *Public Health Genomics*, vol. 15, no. 2, pp. 106–114, 2012.

[33] T. Haeusermann, B. Greshake, A. Blasimme, D. Irdam, M. Richards, and E. Vayena, "Open sharing of genomic data: Who does it and why?," *PLoS ONE*, vol. 12, no. 5, p. e0177158, May 2017.

[34] H. B. Rahmouni, T. Solomonides, M. C. Mont, S. Shiu, and M. Rahmouni, "A Model-driven Privacy Compliance Decision Support for Medical Data Sharing in Europe," *Methods Inf. Med.*, vol. 50, no. 4, pp. 326–336, 2011.

[35] H. B. Rahmouni, T. Solomonides, M. C. Mont, and S. Shiu, "Privacy compliance and enforcement on European healthgrids: an approach through ontology," *Philos. Trans. R. Soc. Lond. Math. Phys. Eng. Sci.*, vol. 368, no. 1926, pp. 4057–4072, Sep. 2010.

[36] B. Blobel, "Ontology driven health information systems architectures enable pHealth for empowered patients," *Int. J. Med. Inf.*, vol. 80, no. 2, pp. e17–e25, Feb. 2011.

[37] F. Falcao-Reis, A. Costa-Pereira, and M. Correia, "Access and privacy rights using web security standards to increase patient empowerment," in *Medical and Care Compunetics 5*, vol. 137, IOS Press, 2008, pp. 275–285.

[38] P. Hung and Y. Zheng, "Privacy Access Control Model for Aggregated e-Health Services," *EDOC Conf. Workshop 2007 EDOC 07 Elev. Int. IEEE*, no. 15-16 Oct. 2007, pp. 12–19, Jul. 2008.

[39] S. Jajodia and D. Wijesekera, "A Flexible Authorization Framework for E-Commerce," in *Distributed Computing and Internet Technology*, 2004, pp. 336–345.

[40] W. M. P. van der Aalst and A. H. M. ter Hofstede, "YAWL: yet another workflow language," *Inf. Syst.*, vol. 30, no. 4, pp. 245–275, Jun. 2005.

www.iariajournals.org

International Journal On Advances in Intelligent Systems
issn: 1942-2679

International Journal On Advances in Internet Technology
issn: 1942-2652

International Journal On Advances in Life Sciences
issn: 1942-2660

International Journal On Advances in Networks and Services
issn: 1942-2644

International Journal On Advances in Security
issn: 1942-2636

International Journal On Advances in Software
issn: 1942-2628

International Journal On Advances in Systems and Measurements
issn: 1942-261x

International Journal On Advances in Telecommunications
issn: 1942-2601