Towards the development of a unified virtual population model in hypertrophic cardiomyopathy

Grigoris I. Grigoriadis
Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, GR 45110 Greece
g.grigoriadis@uoi.gr

Vasileios C. Pezoulas
Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, GR 45110 Greece
v.pezoulas@uoi.gr

Maria Roumpi
Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, GR 45110 Greece
mroumpi89@gmail.com

George Gkois
Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, GR 45110 Greece
gkois@yahoo.com

Nikolaos S. Tachos
Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, GR 45110 Greece
ntachos@gmail.com

Momcilo Prodanovic
Bioengineering Research and Development Center (BioIRC) University of Kragujevac, Faculty of Engineering
Kragujevac, Serbia
ORCID: 0000-0003-0556-1213

Danica Prodanovic
Bioengineering Research and Development Center (BioIRC) University of Kragujevac, Faculty of Science
Kragujevac, Serbia
ORCID: 0000-0002-4770-5468

Boban Stojanovic
Bioengineering Research and Development Center (BioIRC) University of Kragujevac, Faculty of Science
Kragujevac, Serbia
ORCID: 0000-0003-2901-0061

Srboljub M. Mijailovich
Department of Biology Illinois Institute of Technology
Chicago, IL, USA
ORCID: 0000-0003-2640-3548

Nenad Filipovic
BioIRC, Prvoslava Stojanovica 6, Kragujevac, 34000, Serbia
fica@kg.ac.rs

Dimitrios I. Fotiadis
Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH, Ioannina, Greece and the Dept. of Materials Science and Engineering, Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR 45110, Ioannina, Greece
fotiadis@uoi.gr

***Abstract*— The SILICOFCM platform is an in-silico cloud computing platform which utilizes advanced computational workflows for drug development and optimized clinical therapy in the domain of hypertrophic cardiomyopathy (HCM). The current study presents the SILICOFCM's virtual population model (VPM) which can be used to generate high-quality virtual clinical data using both multivariate and machine learning methods along with virtual geometries for in-silico clinical trials. The proposed VPM workflow includes data quality management functionalities for outlier detection and similarity detection which are used to enhance the quality of the real patient data. In addition, the virtual clinical data generator which is part of the VPM includes both multivariate methods, such as, the multivariate normal distribution and machine learning methods, such as, the tree ensembles, the artificial neural networks, and the Bayesian networks. The VPM was utilized in a use-case scenario which included 592 records of patients with HCM towards the generation of clinical data for 1000 virtual patients. Our results suggest that the VPM was able to yield virtual distributions with an increased convergence with the real distributions, where the average goodness of fit was 0.038, the Kullback-Leibler (KL) divergence was 0.029 and the absolute correlation difference 0.0443 between the real and the virtual correlation matrices along with virtual geometries that mimic the real ones.**

***Keywords*— Virtual population, Left ventricle, Virtual geometries, *in-silico* cloud computing platform.**

I. INTRODUCTION

The term *in-silico* clinical trials (ISCT) [1] refers to the utilization of personalized computer simulations for the design, development and evaluation of: (i) medicinal products, (ii) medical devices, and (iii) medical interventions. ISCT is a subdomain of "*in-silico* medicine" and exploits the usage of personalized computer simulations for the disease prevention, diagnosis, prognosis, and treatment. ISCT is an innovative approach for reducing, refining, and partially replacing the real clinical trials [1]. This new approach is already being used for the development of biomedical products. Furthermore, several pharmaceutical companies utilize computer models to estimate or simulate the pharmacokinetics and the pharmacodynamics of a new compound.

The generation of virtual patients could introduce a new virtual treatment enabling the observation, through a computer simulation, regarding the drug performance. There are several state-of-the-art cloud computing platforms which provide advanced functionalities for drug testing in *in-silico* clinical trials. The Simcyp an example of such a platform which

* This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777204 (SILICOFCM).

provides a Population-based Absorption, Distribution, Metabolism and Excretion simulator (ADME) framework [2] for *in-silico* drug research, development, and generation of a physiologically based pharmacokinetic modelling profiles. Moreover, the Nova Discovery Virtual population [3] allows the prediction of the disease progress through the integration of virtual patients and drug candidates with disease models. The platform uses statistical virtual populations composed by specific patient medical parameters, such as the gender, the age, the weight, the height, the systolic and diastolic blood pressure, the total and HDL cholesterol, the glycaemia, the serum creatinine, the left ventricular hypertrophy.

As far as the generation of virtual clinical data is concerned, various parametric methods including the multivariate normal distribution (MVND) and its variant the multivariate log-normal distribution have been deployed in [4, 5] towards the generation of virtual distributions based on real clinical data. In addition, in [6] multinomial logistic models were used to model sequence count data with complex covariance structure. In [7] Bayesian networks were used for virtual population generation by taking into consideration the conditional probabilities among the features and in [8] advanced machine learning methods, such as, the tree ensembles and the artificial neural networks were developed to yield high-quality virtual data.

None of the above studies nor platforms have been able to integrate high-quality virtual clinical data with virtual patient geometries into a unified virtual population model (VPM). In this work, we present the SILICOFCM cloud computing platform which utilizes a set of computational tools towards the generation of virtual patient data including virtual clinical data and heart geometries to deal with the lack of virtual patient models for *in-silico* clinical trials in the domain of hypertrophic cardiomyopathy (HCM). The virtual population model (VPM) is presented and validated towards the generation of virtual clinical data and 3D virtual geometries of patients with HCM, yielding high-quality virtual distributions with goodness of fit 0.038 and Kullback-Leibler (KL) divergence 0.029, as well as synthetic geometries that mimic the real ones.

II. Materials and Methods

A. An overview of the workflow

The proposed workflow for the generation of the virtual patient model (VPM) in HCM is depicted in Fig. 1. The workflow consists of the following modules: (i) the data quality management module, which includes functionalities for outlier detection, similarity detection and data quality reporting, (ii) the virtual clinical data generation module which provides both multivariate and machine learning methods for the generation and validation of the virtual clinical data, and (iii) the virtual biventricular heart geometries generation module which includes functionalities for the parametric heart geometry (left-ventricle and biventricular) generation. The output of the workflow includes: (i) high-quality virtual generated clinical data, (iii) virtual generated ideal LV (left ventricle)/bi-ventricular heart geometries, and (iii) 3D reconstructed patient specific heart geometries.

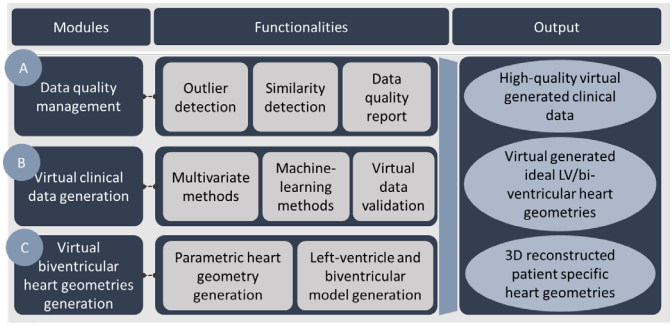

Fig. 1: An illustration of the VPM workflow.

B. Data quality management module

A data curation pipeline which was presented in previous studies [8–10] was utilized for the precise detection of data recording errors including outliers, inconsistencies, and duplicated fields. A quality indicator was developed based on the presence of the previous types of errors according to which the features were categorized into three types, namely the "good", "fair", and "bad" features. Univariate methods, such as, the interquartile range (IQR) [8–10], the z-score [8–10], and the Grubb's test [10] were used to detect outliers based on statistical dispersion. Similarity detection was applied to detect duplicated features [8, 9] by computing: (i) the Spearman rank order correlation coefficient [8, 9] to detect highly associated features, and (ii) the Levenshtein distance [8, 9] to detect lexically matched features.

C. Virtual clinical data generation module

1) Virtual clinical data generation

Both multivariate and machine learning based virtual data generators are applied on the curated data to enhance the quality of the generated virtual clinical data. The cloud-based platform supports six computational methods for the generation of high-quality virtual data (in terms of increased convergence with the real data) for *in-silico* cardiomyopathies clinical trials, as described in [8, 11]. The multivariate methods include: (i) the multivariate normal distribution (MVND) which applies multi-dimensional normal distributions given the mean vector and the covariance matrix of the data, and (ii) the log-MVND which strengthens the assumption of the normality in the MVND through a logarithmic transformation. The machine learning-based methods include: (i) the supervised tree ensembles, which build an ensemble similar to the random forests (RFs) that uses empirical distribution functions to generate virtual data, (ii) the unsupervised tree ensembles, where the ensemble is built in a similar way but instead of random forests, this generator builds a density forest ensemble, (iii) artificial neural networks (ANNs) using RBF (radial basis function) based kernels as activation functions, and (iv) the Bayesian networks, where the structure of the network is used to generate new values with causal dependencies between the features.

2) Virtual clinical data validation

To evaluate the level of agreement of the virtual data in terms of increased convergence and similarity with the real data, three performance evaluation measures were employed from the literature including [8, 11]: (i) the goodness of fit

which measures the distance of two distributions, (ii) the Kullback-Leibler (KL) divergence which quantifies the divergence between two distributions, and (iii) the correlation coefficient which measures the similarity between two distributions.

D. Virtual biventricular heart*geometries generation module*

1) Parametric geometry generation

A straightforward algorithmic approach was implemented, to develop a tool, able to generate 3D biventricular geometries. The developed tool is capable to generate and perform finite elements analysis (FEA), parametric modeling of flat and curved shapes (such as SPLINES, Bezier curves) and parametric geometric surfaces. The VPM tool was developed for creating specific 3D ventricular heart models.

The user is able to define the following ventricular input parameters: (i) the radius of the apex, (ii) the height, (iii) the wall thickness and (iv) the center (x, y, z) coordinates in a .txt file (Fig 2) and obtain the generated 3D model as an output standard tessellation language (STL) file.

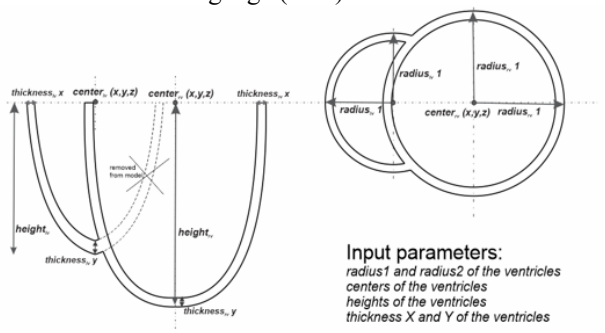

Fig 2: Input parameters for biventricular geometry.

2) Left ventricle and biverntricular model generation

The developed tool provides two options for generating idealized heart geometry: (i) left ventricle model and (ii) biventricular model. Using the parametric modeling approach, the generation of the 3D geometries is fulfilled with outer, inner and top surfaces. Afterwards, the resulting 3D model can be further utilized for mesh generation the finite element analysis (FEA) [12, 13]. The key problem is the dimensional and geometrical variations among each individual patient characteristics. The developed tool which is part of SILICOFCM VPM provides the capabilities to the end-user to generate 3D heart models. The user has just to define the necessary parameters to the input file or to the platform GUI.

3) Geometry morphing

The SILICOFCM VPM, enables an additional option to the end-user for morphing the patient specific geometries to generate virtual. The real heart model is preloaded and the points on the model are prescribed along with displacements at these points. The developed model constrains specific parts of the atlas geometry so that the model does not deform in any direction. Based on the prescribed displacements and constrains provided by the end user, the module deforms the model using polyharmonic deformations [14]. This mesh-based shape deformation method satisfies user deformation constraints at handles (selected vertices or regions on the mesh) and propagates these handle deformations to the rest of shape smoothly and without removing or distorting the details.

III. Results

A. Virtual geometries

An example of the generated 3D geometries is depicted in Fig. 3, where an example including left ventricular and biventricular geometries is depicted.

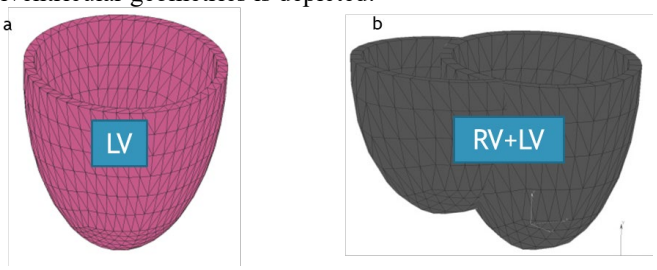

Fig. 3: Results of ventricular generator, a) Left ventricle model, b) biventricular model.

The results from the generated geometries were tested through the implementation of some realistic scenarios. Particularly, a use case scenario where the lateral thickening of the biventricular wall is performed. A point is selected on the heart wall and the user defines the desired displacement. Here the ventricular surfaces are constrained in order not to be displaced. Fig. 4 depicts the lateral thickening of the biventricular wall.

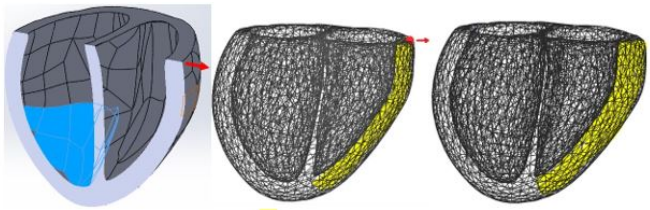

Fig. 4: Use case - lateral wall thickening.

B. Virtual clinical data

The SICLICOFCM database currently includes 592 records (365 retrospective, 229 prospective) of anonymized patient data. The available data include more than 100 features related to demographic (e.g., age, gender), laboratory measures (e.g., systolic pressure), and gene-related information (e.g., CSRP3). The database was utilized for the application of the virtual clinical data generator towards the generation of 1000 virtual patients using the unsupervised tree ensembles algorithm, where the set of real features included the: age, sex, BMI (body mass index), NYHA (New York Heart Association) functional classification, systolic and diastolic pressures, LVIDs (left ventricular internal dimension end-systolic), weight, and height. According to Fig. 5, the level of agreement between the real (blue color) and the virtual data (green color) was favorable, yielding average gof 0.038 and KL divergence 0.029.

According to Fig. 6, the increased convergence between the real with the virtual distributions is validated by the low average absolute correlation difference between the real and the virtual correlation matrix which was 0.0443 ($\pm$0.0319).

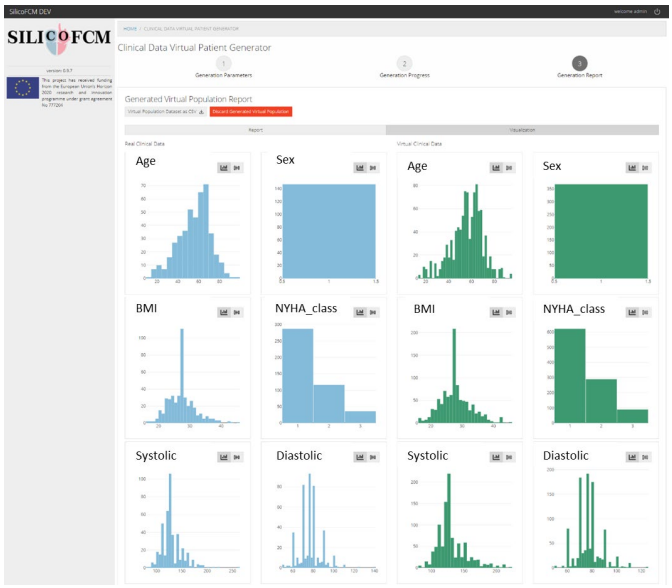

Fig. 5: An instance of the platform's UI with the results of the virtually generated distributions.

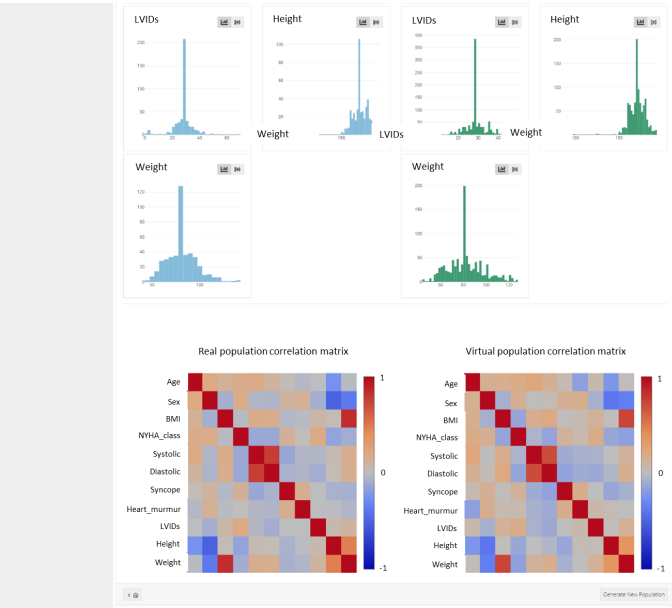

Fig. 6: A second instance of the platform's UI with the rest of the results from the virtually generated distributions along with the correlation matrices of the real and the virtual data.

IV. Discussion And Conclusions

The VPM includes virtual generated clinical data and virtual generated ideal LV/bi-ventricle heart geometries. Specifically, the VPM includes an automatic state-of-the-art data assessment and curation workflow which enhances the pooling of data on the SILICOFCM clinical data repository promoting the high quality of the platform data which is the cornerstone of high-quality simulations and *in-silico* clinical trials through the various tools.

The SILICOFCM VPM supports the virtual generation of high-quality virtual clinical data related to cardiomyopathy and provides a module that is beyond the state, since it provides integrated and advanced data generators. The generated virtual population and the heart3D geometries will be used for further analysis using computational tools. The patient specific 3D reconstructed geometries repository from HCM patients, can be implemented in the 3D morphing pipeline to deform specific geometrical regions generate various HCM case scenarios. Summarizing, the SILICOFCM VPM is an innovative approach for creating a pool of diverse virtual plausible patients' representative of the real cardiomyopathy target population. To the best of our knowledge this innovative and beyond the state-of-the-art *in-silico* solution is the only one available in the literature paving the way for the conceptualization and realization of *in-silico* clinical trials for cardiomyopathies.

References

[1] M. Viceconti, A. Henney, and E. Morley-Fletcher, "In silico clinical trials: how computer simulation will transform the biomedical industry," *Int. J. Clin. Trials*, vol. 3, no. 2, p. 37, May 2016, doi: 10.18203/2349-3259.ijct20161408.

[2] M. Jamei, S. Marciniak, K. Feng, A. Barnett, G. Tucker, and A. Rostami-Hodjegan, "The Simcyp ® Population-based ADME Simulator," *Expert Opin. Drug Metab. Toxicol.*, vol. 5, no. 2, pp. 211–223, Feb. 2009, doi: 10.1517/17425250802691074.

[3] J.-P. Boissel, C. Auffray, D. Noble, L. Hood, and F.-H. Boissel, "Bridging Systems Medicine and Patient Needs: Systems Medicine and Patient Needs," *CPT Pharmacomet. Syst. Pharmacol.*, vol. 4, no. 3, pp. 135–145, Mar. 2015, doi: 10.1002/psp4.26.

[4] D. Teutonico *et al.*, "Generating Virtual Patients by Multivariate and Discrete Re-Sampling Techniques," *Pharm. Res.*, vol. 32, no. 10, pp. 3228–3237, Oct. 2015, doi: 10.1007/s11095-015-1699-x.

[5] S. J. Tannenbaum, N. H. G. Holford, H. Lee, C. C. Peck, and D. R. Mould, "Simulation of Correlated Continuous and Categorical Variables using a Single Multivariate Distribution," *J. Pharmacokinet. Pharmacodyn.*, vol. 33, no. 6, pp. 773–794, Nov. 2006, doi: 10.1007/s10928-006-9033-1.

[6] R. Allen, T. Rieger, and C. Musante, "Efficient Generation and Selection of Virtual Populations in Quantitative Systems Pharmacology Models: Generation and Selection of Virtual Populations," *CPT Pharmacomet. Syst. Pharmacol.*, vol. 5, no. 3, pp. 140–146, Mar. 2016, doi: 10.1002/psp4.12063.

[7] J. D. Silverman, K. Roche, Z. C. Holmes, L. A. David, and S. Mukherjee, "Bayesian Multinomial Logistic Normal Models through Marginally Latent Matrix-T Processes," *ArXiv190311695 Stat*, Apr. 2019, Accessed: Apr. 18, 2021. [Online]. Available: http://arxiv.org/abs/1903.11695.

[8] V. C. Pezoulas, G. I. Grigoriadis, N. S. Tachos, F. Barlocco, I. Olivotto, and D. I. Fotiadis, "Generation of virtual patient data for in-silico cardiomyopathies drug development using tree ensembles: a comparative study," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, Jul. 2020, pp. 5343–5346, doi: 10.1109/EMBC44109.2020.9176567.

[9] V. C. Pezoulas *et al.*, "Medical data quality assessment: On the development of an automated framework for medical data curation," *Comput. Biol. Med.*, vol. 107, pp. 270–283, Apr. 2019, doi: 10.1016/j.compbiomed.2019.03.001.

[10] V. C. Pezoulas, T. P. Exarchos, and D. I. Fotiadis, *Medical data sharing, harmonization and analytics*, 1st ed. San Diego: Academic Press is an Imprint of Elsevier, 2020.

[11] V. Pezoulas, N. Tachos, and D. Fotiadis, "Generation of Virtual Patients for in Silico Cardiomyopathies Drug Development," in *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, Athens, Greece, Oct. 2019, pp. 671–674, doi: 10.1109/BIBE.2019.00126.

[12] H. Si, "TetGen, a Delaunay-Based Quality Tetrahedral Mesh Generator," *ACM Trans. Math. Softw.*, vol. 41, no. 2, p. 11:1-11:36, Feb. 2015, doi: 10.1145/2629697.

[13] D. Z. Milasinovic, A. M. Vukicevic, and N. D. Filipovic, "dfemtoolz: An open-source C++ framework for efficient imposition of material and boundary conditions in finite element biomedical simulations," *Comput. Phys. Commun.*, vol. 249, p. 106996, Apr. 2020, doi: 10.1016/j.cpc.2019.106996.

[14] R. Kountchev, S. Patnaik, J. Shi, and M. N. Favorskaya, Eds., *Advances in 3D Image and Graphics Representation, Analysis, Computing and Information Technology: Algorithms and Applications, Proceedings of IC3DIT 2019, Volume 2*, vol. 180. Singapore: Springer Singapore, 2020.