Predictive Risk Analysis for Liver Transplant Patients — eHealth Model Under National Liver Transplant Program, Uruguay

Parag Chatterjee[1], Ofelia Noceti[2], Josemaría Menéndez[3], Solange Gerona[4], Natalia Harguindeguy[5], Melina Toribio[6], Leandro J. Cymberknop[7], Ricardo L. Armentano[8]

[1,6,8] University of the Republic *(Universidad de la República)*, Uruguay

[2,3,4] Military Hospital *(Dirección Nacional de Sanidad de la Fuerzas Armadas)*, Montevideo, Uruguay

[5] University of Montevideo *(Universidad de Montevideo),* Montevideo, Uruguay

[1,7,8] National Technological University *(Universidad Tecnológica Nacional)*, Buenos Aires, Argentina

[1]Corresponding Email: paragc@ieee.org
[1]ORCID: 0000-0001-6760-4704

ABSTRACT — **Recent years have seen a phenomenal change in healthcare paradigms and data analytics clubbed with computational intelligence has been a key player in this field. One of the main objectives of incorporating computational intelligence in healthcare analytics is to obtain better insights about the patients and proffer more efficient treatment. This work is based on liver transplant patients under the National Liver Transplant Program of Uruguay, considering in detail the health parameters of the patients. Applying computational intelligence helped to separate the cohort into clusters, thereby facilitating the efficient risk-group analysis of the patients assessed under the liver transplantation program with respect to their corresponding health parameters, in a predictive pre-transplant perspective. Also, this marks the foundation of Clinical Decision Support Systems in liver transplantation, which act as an assistive tool for the medical personnel in getting a deeper insight to patient health data and thanks to the holistic visualization of the healthcare scenario, also help in choosing a more efficient and personalized treatment strategy.**

KEYWORDS — Healthcare, predictive analytics, decision support system, liver transplant, data analytics, prediction, risk

I. INTRODUCTION

Healthcare is one of the most important areas where data analytics finds its applications and thanks to the recent boost in paradigms like IoT, eHealth and medical informatics, analyzing health-data has reached new heights. Minimizing the response time in diagnosis and treatment is a crucial component in efficient healthcare services, which makes the power of data analytics relevant for faster analysis and use of intelligent methods for better diagnosis [1]. Detailed analysis of health-data expedites the automated diagnosis on one hand, also leading to personalized treatment. On the other hand, it provides the comprehensive and holistic information of a large group of people under treatment. This is fairly advantageous to automate the process of monitoring along with prediction of health-risks obtained from the analysis of the health-data of the patients.

Liver transplantation is the last therapeutic option in patients with end-stage liver disease. It is a high-cost healthcare process, requiring the expertise of a specialized interdisciplinary team, and a close monitoring of patients throughout the entire timeline of their lives. This process generates a large volume of information and data, of a multiple nature. The adequate clinical management of transplant patients impacts their vital prognosis, and decisions on many occasions are made from the interaction of multiple variables.

The domain of healthcare has always been one of the main sectors motivated to adopt new technologies, given that the primary objective is to provide better and more efficient treatment to the patients. The delivery of health care is a complex endeavor at both individual and population levels. A large pool of historical health-data of a person is a huge plus for any medical personnel before starting a thorough treatment [5]. At the clinical level, the tailored provision of care to individuals is guided, in part, by medical history, examination,

vital signs and evidence. In the 21st century, these traditional tenets have been supplemented by a focus on learning, metrics and quality improvement. The collection and analysis of data of good quality are critical to improvements in the effectiveness and efficiency of health-care delivery [2]. The field of computational intelligence applies to a wide range of disciplines in medicine, however in transplantation, it is still a scarcely explored area.

The main objective of this work was to analyze the main determinants of morbidity and mortality in patients registered under the framework of National Liver Transplant Program in Uruguay, using computational intelligence. This deals with analyzing the probability of developing an inherent risk of disease or complication during the entire timeline, which impacts in patient morbi-mortality. In this stage, focus was put on the analysis of cardiovascular risk (CVR). This work thus spotlights on the applied aspects of data analytics, specifically to analyze the health-data of liver transplant patients of the National Center for Liver Transplantation and Liver Diseases, Uruguay intended to determine and predict the risks with the intention of proffering better diagnosis and treatment to the patients.

II. BACKGROUND OF THE COHORT

This study has been performed on the health-data of patients registered under the National Liver Transplantation Program in Montevideo, Uruguay. The patients considered in this study were registered into the program between the years 2014 and 2017 and were evaluated at the time of their registration (pre-transplantation). Based on the assessment of their severity of their illness and comorbidities, patients are qualified for being enlisted for liver transplantation or rejected. From the pool of patients enlisted, some proceed to liver transplantation, whereas some drop out of the list due to progression illness and the rest continue in the waitlist. To arrange and prioritize the patients in the waitlist, the Model for End-Stage Liver Disease (MELD) score is used, in accordance to most of the centers in Latin America. 104 patients consist of the cohort considered here, as the patients demonstrated a high variance in their health-data (Table I). The mean age of the population was 47 years (ranging from 14 years to 70 years), with almost equal number of males and females. The most frequent indication for liver transplantation was cirrhosis, followed by hepatocellular carcinoma and acute liver failure.

TABLE I.   DETAILS OF HEALTH PARAMETERS OF THE COHORT

Parameters	Cohort Properties
Total number of patients	104
Age at the time of evaluation (years)	47 ± 15
Gender	Male: 51% | Female: 49%
BMI (kg/m²)	27 ± 5
Systolic blood pressure (mm Hg)	117 ± 12
Diastolic blood pressure (mm Hg)	67 ± 8
Total Cholesterol (mmol/L)	159 ± 93
Triglycerides	106 ± 71

HDL (mmol/L)	38 ± 24
LDL (mmol/L)	97 ± 60
Total Cholesterol/HDL	12 ± 26
Platelets (x1000)	125 ± 77
Lymphocytes	1333 ± 854
Neutrophils	3722 ± 2040
Monocytes	599 ± 343
Eosinophils	203 ± 249
Basophils	21 ± 45
Glycemia	100 ± 40
Smoking	Yes: 25% | No: 75%
Diabetes	Yes: 23% | No: 77%
Hypertension	Yes: 28% | No: 72%

Values are expressed as Mean ± Standard Deviation / Ratio / Percentage

The aspect of risks constitutes an entire domain. Some the usual risk scores in this case were dependent on the health parameters. For example, Framingham risk for cardiovascular diseases (FR) was taken into account as one of the interesting risk scores. It mostly considered the parameters of age, gender, blood pressure, cholesterol, smoking and diabetes. Another risk score considered was MELD, which took into account dialysis information and parameters like creatinine, bilirubin, INR and sodium. Apart from that, parameters like death (dead/alive), transplant (transplanted / waitlisted) were also considered interesting for analysis as dependent variables for this study.

III. HEALTH-DATA ANALYTICS — REVIEW AND METHODS

The original focus of this study was to analyze deeply all the health-parameters of the cohort and obtain interesting relationships and correlations among the health parameters and the risk scores.

One of the most crucial tasks in this respect is to identify the most suitable data model from measurements of the system inputs and outputs. Especially in the field of diseases and risk prediction, the data handled is mostly multidimensional [3]. Similarly, in this case of the cohort from the transplantation program, the dataset has more than 20 dimensions. Several methods were analyzed in the beginning, before actually choosing one. In the initial phase of data pre-processing and data organization, principal component analysis (PCA) was considered to be used as a dimensionality reduction technique. But though the original intention was to take advantage of the main benefit to PCA, that is to reduce the size of the feature vectors for computational efficiency, it results in loss of important information, especially taking into account several health-parameters have quite small impact on the risk scores, but still stands interesting from a medical point of view. Also, from the perspective of computational intelligence, methods like support vector machine (SVM), artificial neural networks (ANN) were not considered viable in this specific case because of their complexity and training times and their background of supervised learning [9].

In this respect, unsupervised learning techniques were of key interest, since the intention was to analyze

different groups among the cohort and decipher interesting insights. A considerable part of data in such cases often arriving unlabeled, unsupervised learning methods help in finding patterns in the data or to analyze the health-scenario over a big population. In this aspect, clustering techniques like k-means often stand very useful in separating a group of patients into different clusters and then to analyze in detail the salient features and distinct characteristics [3,4]. Especially when the relationships and impact of different health parameters are not known well, clustering techniques are often used to separate a patient population in order to study the influencing factors. In this aspect, k-means clustering count useful because of its simplicity and was chosen to be applied in this study of the cohort.

A. Clustering — All Patients, Individual Parameters

Firstly, the key objective was to analyze the impact of each individual health parameter on the risk scores, for example, FR. From the perspective of different risk groups in the patients. With that intention, clustering was performed the cohort, considering each of the parameters against FR. Despite in some parameters, the impact was obvious (since the parameters were itself a part of FR), in other parameters also, the clustering separated the patient population into distinct clusters. In fact, another parameter was introduced here — vascular age, to relate with the FR.
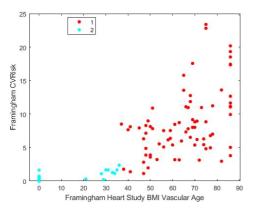

Fig. 1. Vascular Age (using BMI) vs. FR

This showed significant clustering in both cases, when the vascular age was calculated using BMI (Fig. 1) and lipids (Fig. 2).
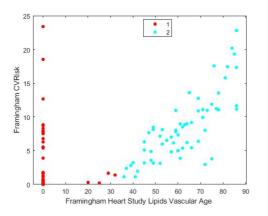

Fig. 2. Vascular Age (using lipids) vs. FR

Nevertheless, it was difficult to analyze the clusters when performed with respect to other parameters not included in the calculation of FR, for example, lymphocytes (Fig. 3), monocytes, neutrophils etc.
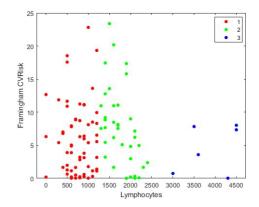

Fig. 3. Lymphocytes vs. FR

This observation lead to the next step, to consider all the parameters while performing the clustering, since the intention would be to analyze the patient risks dependent on the whole health profile instead of just a single parameter.

B. Clustering — All Patients, All Parameters

So aiming at analyzing the full patient health profile, clustering was performed on the entire cohort taking all the features into account. Data cleaning was performed to fix missing values for some patients. The important point is that, the risk parameters like FR, MELD-Na and dependent parameters like death and transplantation condition were also included in this clustering.
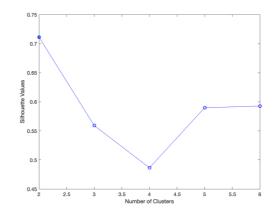

Fig. 4. Silhouette criterion (Optimum Number of Clusters = 2)

The resultant two clusters, chosen based on the silhouette criterion (Fig. 4) indicating that the optimum number of clusters is 2, showed the following characteristics (Table II).

TABLE II.   CLUSTER CHARACTERISTICS (ALL PATIENTS, ALL PARAMETERS)

Parameters	Cluster 1	Cluster 2
Total number of patients	72	25
Age at the moment of evaluation (years)	46 ± 16	46 ± 16
Gender (%)	Male: 54% Female: 46%	Male: 48% Female: 52%
BMI (kg/m^2)	27 ± 5	26 ± 5

Systolic blood pressure (mm Hg)	117 ± 11	116 ± 14
Diastolic blood pressure (mm Hg)	67 ± 8	68 ± 8
Total Cholesterol (mmol/L)	150 ± 54	184 ± 162
Triglycerides	92 ± 42	149 ± 115
HDL (mmol/L)	42 ± 24	29 ± 24
LDL (mmol/L)	90 ± 44	113 ± 92
Total Cholesterol/HDL	7 ± 12	26 ± 47
Platelets (x1000)	107 ± 58	174 ± 97
Lymphocytes	1225 ± 799	1568 ± 908
Neutrophils	2794 ± 1067	6460 ± 1688
Monocytes	524 ± 287	808 ± 413
Eosinophils	194 ± 191	240 ± 382
Basophils	18 ± 42	28 ± 54
Glycemia	98 ± 39	103 ± 45
Framingham Risk	6 ± 5	8 ± 7
MELD	16 ± 6	20 ± 9
Death	Dead: 14%	Dead: 32%
Transplantation	Transplanted: 58%	Transplanted: 68%

Values are expressed as Mean ± Standard Deviation / Ratio / Percentage

Age at the moment of evaluation (years)	46 ± 16	48 ± 14
Gender (%)	Male: 49% Female: 51%	Male: 55% Female: 45%
BMI (kg/m^2)	27 ± 5	27 ± 5
Systolic blood pressure (mm Hg)	117 ± 12	116 ± 12
Diastolic blood pressure (mm Hg)	67 ± 8	68 ± 8
Total Cholesterol (mmol/L)	148 ± 58	175 ± 136
Triglycerides	91 ± 40	134 ± 102
HDL (mmol/L)	42 ± 24	31 ± 24
LDL (mmol/L)	88 ± 45	109 ± 79
Total Cholesterol/HDL	9 ± 18	20 ± 40
Platelets (x1000)	102 ± 59	165 ± 86
Lymphocytes	1100 ± 679	1667 ± 961
Neutrophils	2616 ± 1237	5808 ± 1735
Monocytes	475 ± 274	825 ± 383
Eosinophils	187 ± 202	233 ± 323
Basophils	11 ± 32	36 ± 59
Glycemia	100 ± 40	97 ± 42

Values are expressed as Mean ± Standard Deviation / Ratio / Percentage

The total number of patients considered in this study included patients who have later underwent liver transplants, as well as patients in the waiting list for liver transplantation (Table II). Among the groups were both alive and dead patients who were considered to study their respective physiological characteristics based on some parameters. The key intention was to analyze the differences in trends of the physiological parameters of the groups and also to study their respective risks. Though the two clusters showed quite interesting characteristics, it included all the dependent variables of interest for this analysis, for example, FR, MELD, death, transplantation. It was intuitively evident that these scores also play a significant role in the clustering in the aspect of the separation of the cohort into two clusters.

C. Clustering — All Patients, Without 'Parameters of Interest'

To avoid the impact of the risk parameters in the clustering and to analyze the clusters from an independent and less biased point of view, four parameters (FR, MELD, Death, Transplantation) were not included in the clustering process. This clustering also separated the cohort into two clusters with the following properties (Table III).

TABLE III. CLUSTER CHARACTERISTICS (ALL PATIENTS, WITHOUT 'PARAMETERS OF INTEREST')

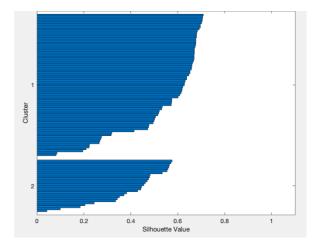| Parameters	Cluster 1	Cluster 2
Total number of patients	61	36

Fig. 5. Silhouette (Clusters with All Patients, All Parameters except 'Parameters of Interest')

The silhouette plot (Fig. 4) using the Euclidean distance metric showed that the data is split into two clusters, one smaller and one bigger, similar to the clusters obtained with all patients and all parameters. Most of the points in the two clusters had large silhouette values (>=0.5), indicating that the clusters were separated quite well. However, since the 4 parameters were not considered in the clustering process, after the clusters were obtained, those 4 parameters were linked to the corresponding patients and their respective characteristics were calculated (Table IV).

TABLE IV. POST-CLUSTERING ANALYSIS (ALL PATIENTS — FR, MELD, DEATH, TRANSPLANTATION)

Parameters	Cluster 1	Cluster 2

Framingham Risk	6 ± 5	7 ± 6
MELD	16 ± 6	19 ± 8
Death	Dead: 16%	Dead: 25%
Transplantation	Transplanted: 54%	Transplanted: 69%

D. Clustering — Transplant and WaitList Patients, All Parameters, without the 4 Parameters of Interest

A similar approach was followed to perform clustering separately on the waitlisted patients and the transplanted patients, but without including FR, MELD-Na, death and transplantation condition in the clustering process. However, in the same manner, those were analyzed after the clustering process with the corresponding patients. In each case of waitlisted and transplanted patients, 2 clusters were obtained with different characteristics. Silhouette analysis was performed in each case (transplant patients and waitlist patients, all parameters except the 4 parameters of interest), and the cluster points had quite large silhouette values (>=0.5), indicating that the clusters were separated quite well. Also, the 4 risk parameters of interest were analyzed in relation to the first two cluster-analyses performed on all the patients.

IV. RESULTS AND DISCUSSIONS

The general observation from the clusters obtained after the entire cluster analysis performed on various instances, every time two clusters were obtained, containing always a bigger cluster with more number of patients and a smaller cluster with lesser number of patients.

In the first study, where clustering was performed on all patients considering all parameters, the smaller cluster showed higher mean value of FR (8) and MELD (20) than the bigger cluster (FR: 6, MELD: 16). Also the smaller cluster showed higher percentage of death (32%) and higher percentage of transplant patients (68%) than the bigger cluster (death: 14%, transplantation: 58%). Apart from that, though both the clusters have almost similar mean age, BMI, and blood pressure, the smaller cluster showed significantly lower value of HDL and higher values for LDL, triglycerides, total cholesterol/HDL, platelets, lymphocytes, neutrophils, monocytes, eosinophils, basophils and glycemia, implying the smaller cluster at higher risk than its bigger counterpart.

In the second case, when the clustering was performed on all the patients but without considering the FR, MELD, death and transplantation, two clusters were obtained, one bigger (61 patients) and the other smaller (36 patients). Here also, the smaller cluster showed higher values of of FR (7), MELD (19), and higher percentage of death (25%) and transplant patients (69%) than the bigger cluster (FR: 6, MELD: 16, death: 16%, transplantation: 54%). Also, though both the clusters have almost similar mean age, BMI, and blood pressure, the smaller cluster showed significantly lower value of HDL and higher values for LDL, triglycerides, total cholesterol/HDL, platelets, lymphocytes, neutrophils, monocytes, eosinophils, basophils and glycemia, implying the smaller cluster at higher risk than the bigger one.

The similar trend followed in the clusters obtained from the waitlisted patients and the transplant patients. In all the cases, apart from the risk scores (FR, MELD) and parameters like death and transplantations, other parameters like lymphocytes, neutrophils, monocytes showed a significant (p<0.05) rise in the smaller cluster than the bigger one.

From the medical perspective, higher lymphocyte-count indicates to possibilities of lymphocytosis, frequently associated with chronic infections, inflammations and autoimmune diseases. Also, higher count of monocytes indicates to potential risk of infection and neutrophil and platelet count signals to the inflammation status.

Intuitively, it turned out, that after every clustering, among the two clusters obtained, the smaller cluster demonstrates more risk and more vulnerable patients than the ones in the bigger cluster. Also, it implied that at the point of evaluation in the timeline, with no knowledge of the future events, the patient population could be divided analytically into two clusters, considering just the parameters obtained during evaluation. Even without taking into account the risk scores like FR and MELD, the patient population could be separated into two groups, with one of those showing significantly higher risks than the other. For example, the mean time to transplant from the point of evaluation being 3 months (ranging even up to 3 years), this clustering model could direct a patient to a cluster with high risk or low risk, just using the parameters at the time of evaluation, while enlisting into the system. Thus, it shows the holistic view to the overall health conditions of a group of people, facilitating the identification of high and risk groups [6].

V. CONCLUSION

Liver transplant itself being a risk, it is important to ascertain the suitability to perform the transplantation and this invokes the need of analysis of other possible risks of the patient before taking the decision. Despite the presence of definite risk scores to determine the severity of the health conditions of a patient in the list for transplant, computational intelligence leaves scope to take advantage of all the health parameters evaluated in ascertaining the risks with higher efficiency. This work takes a step in that direction, using the simple aspects of computational intelligence in segregating the patient cohort into risk groups from a predictive point of view, considering the simple health-parameters which is normally evaluated for every enlisted patient. Though based on the volume of the cohort in this work, the inferences couldn't be scaled to a larger population, but still it leaves scope to analyze a larger cohort as well. Decision support systems are interesting components of recent healthcare systems which analyzes data and supports healthcare providers to take clinical decisions [7,8]. This work leads to the design of a predictive clinical decision support system, aiming at automatically classifying the patient population at the evaluation time into high-risk or low-risk, facilitating the aspect of care during the enlisted period.

REFERENCES

[1] L.E. Romero, P. Chatterjee, R. L. Armentano, "An IoT approach for integration of computational intelligence and wearable sensors for Parkinson's disease diagnosis and monitoring," Health and Technology, 6, (3), pp. 167-172, 2016

[2] WHO. Bulletin of the World Health Organization 2015;93:203-208. doi: http://dx.doi.org/10.2471/BLT.14.139022 [Internet]. 2019. Available from: https://www.who.int/bulletin/volumes/93/3/14-139022/en/

[3] P. Chatterjee, L. J. Cymberknop, and R. L. Armentano, "Nonlinear Systems in Healthcare — A Perspective of Intelligent Disease Prediction," in Nonlinear Systems, vol. 2, W. Legnani, Eds. IntechOpen, 2019, DOI: 10.5772/intechopen.88163. Available from: https://www.intechopen.com/online-first/nonlinear-systems-in-healthcare-towards-intelligent-disease-prediction

[4] Healthcare.ai. Step by Step to K-Means Clustering. Data Science Blog. 21 July 2017, healthcare.ai/step-step-k-means-clustering/

[5] Chatterjee P, Cymberknop L, Armentano R. IoT-Based Decision Support System Towards Cardiovascular Diseases. Córdoba, Argentina: SABI; 2017

[6] Chatterjee P, Cymberknop L, Armentano R. IoT-based ehealth toward decision support system for CBRNE events. In: Malizia A, D'Arienzo M, editors. Enhancing CBRNE Safety & Security: Proceedings of the SICC 2017 Conference. Cham: Springer; 2018. DOI: 10.1007/978-3-319-91791-7_21

[7] Chatterjee P, Armentano RL, Cymberknop LJ. Internet of things and decision support system for eHealth-applied to cardiometabolic diseases. In: International Conference on Machine Learning and Data Science (MLDS); Noida. Piscataway: IEEE; 2017. pp. 75-79. DOI: 10.1109/MLDS.2017.22

[8] Chatterjee P, Cymberknop LJ, Armentano RL. IoT-based decision support system for intelligent healthcare—Applied to cardiovascular diseases. In: 7th International Conference on Communication Systems and Network Technologies (CSNT); Nagpur. 2017. pp. 362-366. DOI: 10.1109/CSNT.2017.8418567

[9] Hippolyte T, Adamou M, Blaise N, Pierre C, Olivier M. Linear vs non-linear learning methods—A comparative study for forest above ground biomass, estimation from texture analysis of satellite images. ARIMA Journal. 2014;18:114-131