Chapter 4

DATA ANALYTICS IN BIOMEDICAL ENGINEERING AND HEALTHCARE

Published by ACADEMIC PRESS | An imprint of **ELSEVIER**

https://www.sciencedirect.com/science/article/pii/B9780128193143000045
DOI: 10.1016/B978-0-12-819314-3.00004-5

Machine learning in healthcare toward early risk prediction: A case study of liver transplantation

Parag Chatterjee[a,b], Ofelia Noceti[c], Josemaría Menéndez[c], Solange Gerona[c], Melina Toribio[b], Leandro J. Cymberknop[a], and Ricardo L. Armentano[a,b]

[a]National Technological University (Universidad Tecnológica Nacional), Buenos Aires, Argentina. [b]University of the Republic (Universidad de la República), Montevideo, Uruguay. [c]Military Hospital (Dirección Nacional de Sanidad de las Fuerzas Armadas), Montevideo, Uruguay

1 Introduction

Healthcare services have seen a strong paradigm shift in recent years globally. At the clinical level, personalized care to individuals is usually provided based on medical history, examination, vital signs, and evidence. However, in the recent times, the focus on these traditional tenets is being taken over by the aspects of learning, metrics, and quality improvement [1]. The last decade has seen a global rise in adoption of Electronic Health Records (EHRs) [2–7], catalyzing the increase in the complexity and volume of the health data generated in the process. Apart from the EHR-sourced ordinary patient data, due to the change in treatment paradigms and focus on lifestyle and comprehensive healthcare, new varieties of health-data about medical conditions, lifestyle, underlying genetics, medications, and treatment approaches also showed a paramount rise. Despite the complex nature of new-generation health data, human cognition to analyze and make sense of these humongous data is finite [8]. The traditional medical models of analysis deserve a reengineering for more efficiency, leading to the computer-assisted methods to organize, interpret, and recognize patterns from these data [9, 10]. Efficient collection and accurate analysis of data are critical to improvements in the effectiveness and efficiency of healthcare delivery [1]. In this respect, emerging as a promising field, eHealth addresses multifarious aspects of the healthcare system, like tracking changes in health behavior and prevention and management of chronic diseases [11].

In the recent years, the intrinsic power of data in healthcare started unveiling like never before, leading to the endeavor in making sense of health data in the best possible way, using advanced data analytics and computational intelligence. Especially in the field of healthcare, the aspect of intelligent data analytics is one of the most trending topics worldwide [12]. One of the prime areas where such analyses have been applied is the field of chronic diseases. By 2020, chronic diseases are expected to contribute to 73% of all deaths worldwide and 60% of the global burden of disease. Also, 79% of the deaths attributed to these diseases occur in the developing countries. Four of the most prominent chronic diseases—cardiovascular diseases (CVD), cancer, chronic obstructive pulmonary disease, and type 2 diabetes—are linked by common and preventable biological risk factors, notably high blood pressure, high blood cholesterol, and overweight, and by related major behavioral risk factors. To prevent these major chronic diseases, the actions need to be centered around controlling the key risk factors in a comprehensive and integrated manner [13]. In addition to the chronic diseases, the aspect of intelligent risk prediction and preventive actions count significant even in the domain of transplantations [10, 14].

Moreover due to the concept of context awareness backed by sensor fusion in the environment of smart eHealth systems and IoT, the health data generated and acquired is more comprehensive and detailed. The smart prediction and prevention systems in healthcare usually share some common steps like the collection of health data from sensors or other sources and assimilating the EHRs, followed by analyzing and computing the risks [15]. In the endeavor of taking possible actions to prevent chronic diseases, detecting the diseases at an early stage stands to be the prime challenge. Most of these diseases do not exhibit clearly identifiable signs at the early stage. And here lies the key area of Artificial Intelligence (AI), harvesting the possibility of early detection of these diseases in terms of risk. From the data science perspective, the aspect of health data acts as the key valuable resource. Most importantly, the domain of health data has expanded dramatically over these years. Even the superficially and noninvasively obtained behavioral, physiological, and metabolic health data hold enormous significance. In the domain of early detection and prediction of diseases, the health data possesses a huge potential. Disease prediction led by AI is a multilevel process. It involves the analysis of the intricate details inside the health data, looking for early indications or traces of diseases.

Thanks to the recent boost in paradigms like IoT, eHealth, and medical informatics pertaining to AI, healthcare is one of the most important areas where data analytics finds its applications and analyzing health data has reached new heights [15a]. 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 [16]. 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. In this direction, one of the key areas where prediction of risk is highly crucial is transplantation. Transplantation is itself a complex procedure that makes the consideration of pretransplant predictions of risk an important factor, opening up a broad scope for computational intelligence. Possible insights obtained from the pretransplant health data of the patients by harnessing the power of AI would positively influence the healthcare delivery approach. In this work a case study is presented highlighting the aspects of computational intelligence toward risk prediction in liver transplantation.

Liver transplantation is the last therapeutic option in patients with end-stage liver disease. Being a complex healthcare process, it is related to humongous costs and requires the expertise of a specialized interdisciplinary team along with a close monitoring of patients during the entire timeline. This process generates a large volume of complex, multidimensional data. The adequate clinical management of transplant patients impacts their vital prognosis, and decisions on many occasions are made from the interaction of multiple variables [17]. However, there exists an enormous demand in the domain of prediction in liver transplantation process, ranging from survival till the suitability of transplant [18].

The healthcare sector has emerged as one of the prime areas to adopt new technologies, given that the primary objective is to provide better and more efficient treatment to the patients. Healthcare delivery is a complex aspect at both individual and population levels. From the perspective of data-driven insights for any medical personnel, a large pool of historical health data of the patient is a huge plus, before starting a thorough treatment [19].

At the clinical level the aspect of providing healthcare services is guided mostly by medical history, examination, vital signs, and evidence. Recent times have seen a paradigm-shift of the core traditional approaches toward the supplementary focus on learning, metrics, and quality improvement of the healthcare provided. The collection and analysis of good-quality data are critical to improvements in the effectiveness and efficiency of delivering healthcare services [20]. The field of artificial intelligence applies to a wide range of disciplines in medicine; however, in transplantation, it is still a scarcely explored area.

The AI has started playing an important role in predicting the main determinants of morbidity and mortality in patients, which stands quite significant in the domain of transplantations. This deals with analyzing the probability of developing an inherent risk of disease or complication during the entire timeline. The main objective of this work is to spotlight the applied aspects of data analytics in healthcare and importance of AI in transplantation, illustrated through a case study of liver transplantation at the National Center for Liver Transplantation and Liver Diseases, Uruguay. Also, based on the advantages of AI in transplantation, an AI-based predictive clinical decision support system for transplantations has been proposed for early detection and prediction of risks and proffering better diagnosis and treatment to the patients.

2 Background of the study: Description of cohort

This study is focused on the patients registered under the National Liver Transplantation Program in Montevideo, Uruguay. In this case the patients considered for the study were registered into the program between the years 2014 and 2017 and were evaluated at the time of their registration to the program. Depending on the assessment at the time of registration and assessing the severity of their illness and comorbidities, the patients qualify to continue in the registered waitlist for liver transplantation or step down from the list. Based on further analysis, some patients from the pool of enlisted patients proceed to liver transplantation, whereas some drop out from 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 liver-treatment centers. 104 patients consist of the cohort considered here (Table 1). The mean age of the population was 47 years (ranging from 14 to 70 years), with almost

Table 1 Details of health parameters of the cohort.

Parameters	Cohort properties
Total number of patients	104
Age at the moment of evaluation (years)	47 ± 15
Gender (%)	Male: 51% | Female: 49%
BMI (kg/m^2)	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 ($\times 1000$)	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 $\pm$ Standard Deviation.

equal number of males and females. The most frequent indication for liver transplantation was cirrhosis, followed by hepatocellular carcinoma and acute liver failure.

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 Cardiovascular Disease 10-year risk (FR) was taken into account as one of the interesting risk factors. It mostly considered the parameters of age, gender, blood pressure, LDL and HDL cholesterol, smoking, and diabetes. Another risk score considered was the MELD, which took into account dialysis information and parameters like creatinine, bilirubin, INR, and sodium. Apart from that, parameters like death (dead/alive) and transplant (transplanted/waitlisted) were also considered interesting for analysis as dependent variables for this study [17].

3 Intelligent risk analysis: Aspect of vascular age and cardiometabolic health

The original focus of this study was to analyze all the health parameters of the cohort and obtain interesting relationships and correlations among the health parameters and the risk scores, especially with the intention to analyze the early detection and prediction of risks in a pretransplant scenario.

One of the most crucial tasks in this respect was 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 are mostly multidimensional [10]. Similarly, in this case, the dataset constituted by the cohort from the transplantation program had more than 20 dimensions. Several methods were analyzed in the beginning, before actually choosing one. In the initial phase of data preprocessing 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, which 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, several supervised learning methods were not considered viable in this specific case because of their complexity and training times and their background of supervised learning [21]. 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 [10]. Briefly speaking, k-means clustering aims to find the set of k clusters such that every data point is assigned to the closest center and the sum of the distances of all such assignments is minimized [22]. Especially when the relationships and impact of different health parameters are not known well, clustering techniques are often used to separate a patient population to study the influencing factors. In this aspect, k-means clustering counts useful because of its simplicity and was chosen to be applied in this study of the cohort.

In the initial phase the key objective was to analyze the impact and relevance of each health parameter with respect to the risk scores. Despite that some of the parameters showed good relationships with the risk scores, some parameters were itself included in the calculation of the risk scores; of course, tautology was evident. Also the focal goal was to understand the overall risk scenario of the cohort, considering all the health parameters and relating it with the risk scores. Aiming at analyzing the full patient health profile, clustering was performed on the entire cohort taking all the features into account. To fix the missing values in some features, data cleaning was performed. However, in this case, the risk parameters like FR, MELD, and dependent parameters like death and transplantation events were also included in this clustering. The silhouette criterion indicated that the optimum number of clusters is two (Fig. 1). After performing the clustering with two clusters, each of the clusters illustrated distinguishing characteristics between each other (Table 2).

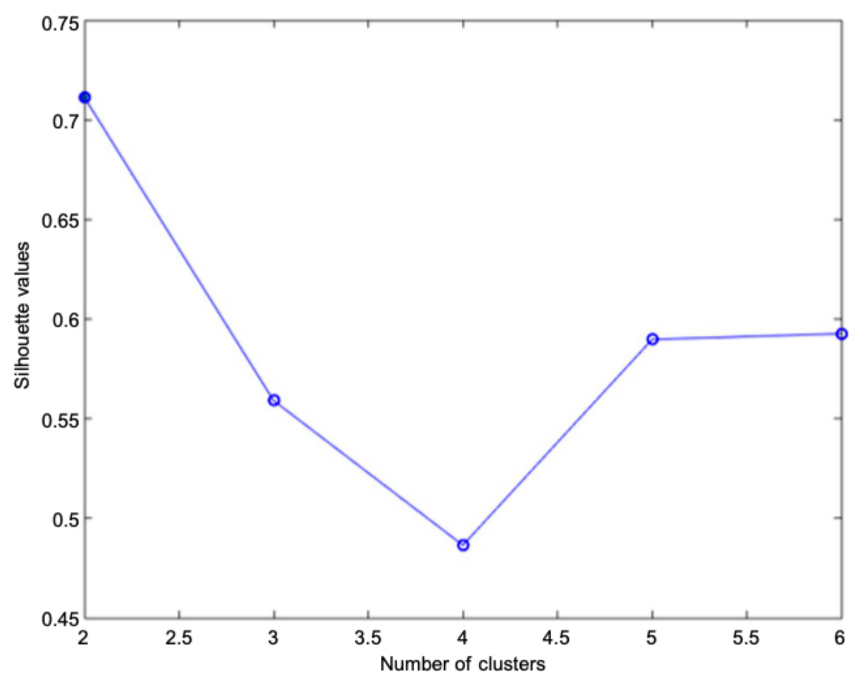

Fig. 1 Silhouette criterion (optimum number of clusters = 2).

Table 2 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 $\pm$ Standard Deviation/Ratio/Percentage.

In the cohort considered in this study, there are patients in the registered list who eventually had liver transplants and patients who continued in the waitlist. Also among each group, the patients were separated into alive and dead to study their respective physiological characteristics. The motive was to assess the trends in physiological parameters of the groups and also to study their distinguishable risks.

Despite the clusters showing quite interesting characteristics, the process itself included several risk scores, or rather dependent variables (e.g., MELD, death, transplantation, and FR). Intuitively, it was evident that these scores also had a significant impact in affecting the separation of the cohort into two clusters. To reduce the impact of the risk scores itself in the clustering process and to

Table 3 Cluster characteristics (all patients, without "parameters of interest").

Parameters	Cluster 1	Cluster 2
Total number of patients	61	36
Age at the moment of evaluation (years)	46 ± 16	48 ± 14
Gender (%)	Male: 49%	Male: 55%
	Female: 51%	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

analyze the clusters from an independent and less biased point of view, four parameters (FR, MELD, death, and transplantation) were not included in the clustering process. This clustering also separated the cohort into two clusters with the following properties (Table 3). The silhouette plot using the Euclidean distance metric showed that the data are 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), indicating that the clusters were separated quite well. However, since the four parameters were not considered in the clustering process, after the clusters were obtained, those four parameters were linked to the corresponding patients, and their respective characteristics were analyzed (Table 4) [17].

A similar approach was followed to perform clustering separately on the waitlisted patients and the transplanted patients,

Table 4 Postclustering analysis (all patients—FR, MELD, death, and 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%

but without including FR, MELD, 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, two clusters were obtained with distinguishable characteristics. Silhouette analysis was performed in each case (transplant patients and waitlist patients, all parameters except the four parameters of interest), and the cluster points indicated that the clusters were separated quite well. Also the four risk parameters of interest were analyzed in relation to the first two cluster-analyses performed on all the patients [17].

In an extended version of this study, the aspect of risks has been widened to the domain of cardiovascular health. Human body being a comprehensive and complex structure, the important relationship between the liver and the circulatory system is quite evident. The latter, made up of the heart and all the body's blood vessels, is affected by various parameters that can impair or improve its functioning and, therefore, the health of the individual. This leads to focusing on the state of the patients' cardiovascular health, to have a snapshot of the relationship between the parameters that interfere with it and the health risks of liver patients as well. In this respect the Framingham heart study was used as the model to analyze the cardiovascular risks of the cohort. The 10-year cardiovascular disease risk function [23] was used, considering the parameters like age, gender, total cholesterol, HDL, hypertension, diabetes, and smoking, which in turn returned the Framingham Risk score, along with the vascular age of the respective patient. The percentage of risk (10-year prediction) considered here reflects the possibility of suffering a cardiovascular event like coronary death, myocardial infarction, coronary insufficiency, angina, ischemic stroke, hemorrhagic stroke, transient ischemic attack, peripheral arterial disease, or heart failure. Specifically in this case an extended cohort of 170 patients evaluated between 2014 and 2019 were considered, where 55% of patients were transplanted, and the remaining 45% were divided into waitlisted patients, who were removed from the list due to improved health, or due to death after the evaluation and before receiving the transplant. In this cohort the risk function was applied, and Framingham risk score and vascular age for the patients were calculated. Also, for each patient, the difference between their vascular age and biological age was calculated. Based on the ΔAge (Vascular Age – Biological Age), the cohort was separated into two groups, where $\Delta\text{Age} > 0$ (Vascular Age > Biological Age) and $\Delta\text{Age} \leq 0$ (Vascular Age ≤ Biological Age).

4 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 and a smaller cluster.

In the first study, where clustering was performed on the all patients and 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 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 and MELD) and parameters like death and transplantations, other parameters like lymphocytes, neutrophils, and 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 successfully into two clusters, considering just the parameters obtained during evaluation. Even without taking into account the risk parameters 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 (the maximum being 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. Such approach is apt for inclusion in clinical decision support systems in transplantations. A typical non–knowledge based clinical decision support system (CDSS) holds in its core the power of machine learning. Feeding the comprehensive health records of the patient cohort into the CDSS makes it possible for artificial intelligence to analyze the cohort and classify them into risk groups, eventually alerting the medical personnel the dynamic possibilities of complications for a particular patient in a pretransplant scenario. Thus, it shows the holistic view to the overall health conditions of a group of people, facilitating the identification of high and risk groups [17, 24].

In the extended study of the cardiometabolic risk groups within the patient cohort, the separated groups based on their ΔAge (Vascular Age − Biological Age) showed distinguishing properties with respect to their health parameters (Table 5).

Analyzing the two groups with respect to their ΔAge (Vascular Age − Biological Age), it is quite natural that the group of patients with $\Delta\text{Age} > 0$ has higher BMI and higher percentage of smoking, diabetes, and hypertension, along with significantly high values of total cholesterol/HDL than the group of patients with $\Delta\text{Age} \leq 0$. This can be explained in terms of cardiometabolic risks due to the presence of those parameters within the calculation of vascular age itself. But in addition to that, the first group with $\Delta\text{Age} > 0$ showed higher number of MELD score and higher percentage of transplantations. The MELD score estimates a patient's chances

Table 5 Risk-group properties (separated cohort in terms of vascular age).

Parameters	Vascular age > biological age	Vascular age ≤ biological age
Total number of patients	75% of the cohort	25% of the cohort
Age at the time of evaluation (years)	48 ± 15	49 ± 12
Gender	Male: 59% | Female: 46%	Male: 39% | Female: 61%
BMI (kg/m^2)	28 ± 6	26 ± 4
Systolic blood pressure (mm Hg)	119 ± 14	106 ± 13
Diastolic blood pressure (mm Hg)	69 ± 10	63 ± 8
Total cholesterol (mmol/L)	159 ± 93	143 ± 49
Triglycerides	117 ± 82	96 ± 67
HDL (mmol/L)	30 ± 21	47 ± 26
LDL (mmol/L)	103 ± 60	78 ± 32
Total cholesterol/HDL	14 ± 27	4 ± 4.4
Platelets ($\times$1000)	119 ± 76	110 ± 54
Lymphocytes	1435 ± 1476	1356 ± 978
Neutrophils	4199 ± 3691	3924 ± 3549
Monocytes	623 ± 453	514 ± 321
Eosinophils	189 ± 239	150 ± 158
Basophils	22 ± 46	8 ± 28
Glycemia	108 ± 48	91 ± 21
Smoking	Yes: 32% | No: 68%	Yes: 3% | No: 97%
Diabetes	Yes: 39% | No: 61%	Yes: 19% | No: 81%
Hypertension	Yes: 34% | No: 66%	Yes: 14% | No: 86%
MELD Score	18 ± 7	16 ± 8
Framingham CV risk score	15 ± 13	5 ± 3
Vascular age	62 ± 19	43 ± 12
Vascular Age–Biological Age	14 ± 9	-7 ± 4
Transplantation	Yes: 62% No: 38%	Yes: 56% No: 44%

of surviving their disease during the next 3 months, and physicians working on the liver transplant program in this case used MELD score to classify the level of liver severity of patients and determine the urgency of the liver transplant [25], with a higher MELD indicating higher severity. It has been observed in this study that the group of patients having their vascular age more than their actual biological age also showed higher severity with respect to MELD score and eventually showed higher percentage of transplantations, indicating a possible relationship of transplantations with cardiometabolic risks.

The study and the analysis invoke the possibility of evaluation of a new patient entering the evaluation list for the liver transplantation program using a CDSS, to predict the cardiometabolic risks along with the usual evaluation procedure and instantaneously assign the risk group the new patient falls in. This would be a key aspect of a potentially assistive tool to the medical personnel to classify the patient cohort into risk groups at any given time starting from the entry-point evaluation.

5 Conclusion

In the new age of data and eHealth, the inherent knowledge of data has turned out to be of immense importance, and computational intelligence plays a key role in making the most out of the data. Especially for chronic diseases, long-term behavioral and lifestyle data stand quite crucial. Data modeling and predictive analytics open a huge avenue toward clinical decision support systems, which is a fundamental tool nowadays for preventive and personalized healthcare and supports healthcare providers to have deeper insights into patients' data [26] and take clinical decisions [27]. Transplantation are associated with several risk factors, which if predicted better using computational intelligence, stands quite paramount in reducing the mortality in transplantation process. In this case study of liver transplantation, it was 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 artificial intelligence in segregating the patient cohort into risk groups from a predictive point of view, considering the simple health parameters, which are normally evaluated for every enlisted patient. Though based on the volume of the cohort in this work, it is difficult to design precise risk-model, but the inferences could be scaled to a larger population leaving the scope to validate in a larger cohort as well. Decision support systems are interesting components of recent healthcare systems, which analyze data and support healthcare providers to take clinical decisions [26, 27]. This work leads to the idea 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 [17].

Acknowledgments

This work is financially supported by the National Agency for Research and Innovation (*Agencia Nacional de Investigación e Innovación*) of Uruguay through its grant FSDA-2017-1-143653 (*Inteligencia Computacional en Salud. Creando Herramientas para Predecir la Sobrevida en Trasplante Hepático*).

References

[1] WHO. Bull. World Health Organ. 2015;93:203–208. https://doi.org/10.2471/BLT.14.139022 [Internet]. 2019. Available from: https://www.who.int/bulletin/volumes/93/3/14-139022/en/.

[2] Charles DK J, Patel V, Furukawa M. Adoption of Electronic Health Record Systems among U.S. Non-federal Acute Care Hospitals: 2008–2012; 2013. http://www.healthit.gov/sites/default/files/oncdatabrief9final.pdf. Accessibility verified April 20, 2014.

[3] N.H. Shah, Translational bioinformatics embraces big data, Yearb. Med. Inform. 7 (1) (2012) 130–134.

[4] O. Heinze, M. Birkle, L. Koster, B. Bergh, Architecture of a consent management suite and integration into IHE-based regional health information networks, BMC Med. Inform. Decis. Mak. 11 (2011) 58.

[5] A. Tejero, I. de la Torre, Advances and current state of the security and privacy in electronic health records: survey from a social perspective, J. Med. Syst. 36 (5) (2012) 3019–3027.

[6] A. Mense, F. Hoheiser-Pfortner, M. Schmid, H. Wahl, Concepts for a standard based cross-organisational information security management system in the context of a nationwide EHR, Stud. Health Technol. Inform. 192 (2013) 548–552.

[7] A. Faxvaag, T.S. Johansen, V. Heimly, L. Melby, A. Grimsmo, Healthcare professionals' experiences with EHR-system access control mechanisms, Stud. Health Technol. Inform. 169 (2011) 601–605.

[8] M.K. Ross, W. Wei, L. Ohno-Machado, "Big data" and the electronic health record, Yearb. Med. Inform. 9 (1) (2014) 97–104, https://doi.org/10.15265/IY-2014-0003.

[9] K.B. Wagholikar, V. Sundararajan, A.W. Deshpande, Modeling paradigms for medical diagnostic decision support: a survey and future directions, J. Med. Syst. 36 (5) (2012) 3029–3049.

[10] P. Chatterjee, L.J. Cymberknop, R.L. Armentano, Nonlinear systems in healthcare—a perspective of intelligent disease prediction, in: W. Legnani (Ed.), Nonlinear Systems, vol. 2, IntechOpen, 2019https://doi.org/10.5772/intechopen.88163https://www.intechopen.com/online-first/nonlinear-systems-in-healthcare-towards-intelligent-disease-prediction.

[11] D.K. Ahern, J.M. Kreslake, J.M. Phalen, What is eHealth (6): perspectives on the evolution of eHealth research. J. Med. Internet Res. 8 (1) (2006) e4, https://doi.org/10.2196/jmir.8.1.e4.

[12] V.E. Balas, S.S. Roy, D. Sharma, P. Samui (Eds.), Handbook of Deep Learning Applications, In: vol. 136, Springer, 2019.

[13] WHO. Integrated Chronic Disease Prevention and Control [Internet]. 2019. Available from: https://www.who.int/chp/about/integrated_cd/en/.

[14] L.B. VanWagner, H. Ning, M. Whitsett, J. Levitsky, S. Uttal, J.T. Wilkins, M. M. Abecassis, D.P. Ladner, A.I. Skaro, D.M. Lloyd-Jones, A point-based prediction model for cardiovascular risk in orthotopic liver transplantation: the CAR-OLT score. Hepatology 66 (2017) 1968–1979, https://doi.org/10.1002/hep.29329.

[15] Masoud Hemmatpour, Renato Ferrero, Filippo Gandino, Bartolomeo Montrucchio, and Maurizio Rebaudengo. Nonlinear predictive threshold model for real-time abnormal gait detection. J. Healthc. Eng., vol. 2018, Article ID 4750104. 9 pages. https://doi.org/10.1155/2018/4750104.

[15a] P. Chatterjee, R. Armentano, L. Palombi, L. Kun, Editorial preface: special issue on IoT for eHealth, elderly and aging. Internet of Things (2019) 100115, https://doi.org/10.1016/j.iot.2019.100115.

[16] 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 Technol. 6 (3) (2016) 167–172, doi: 10.1007/s12553-016-0148-0.

[17] P. Chatterjee, et al., Predictive risk analysis for liver transplant patients—eHealth model under National Liver Transplant Program, Uruguay. IEEE 9th International Conference on Advanced Computing (IACC), Tiruchirappalli, India 2019 (2019) 75–80, https://doi.org/10.1109/IACC48062.2019.8971514.

[18] D. Bertsimas, J. Kung, N. Trichakis, Y. Wang, R. Hirose, P.A. Vagefi, Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation, Am. J. Transplant. 19 (4) (2019) 1109–1118.

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

[20] WHO. Bull. World Health Organ. 2015;93:203–208. doi: https://doi.org/10.2471/BLT.14.139022 [Internet]. 2019. Available from: https://www.who.int/bulletin/volumes/93/3/14-139022/en/.

[21] T. Hippolyte, M. Adamou, N. Blaise, C. Pierre, M. Olivier, Linear vs non-linear learning methods—a comparative study for forest above ground biomass, estimation from texture analysis of satellite images, ARIMA J. 18 (2014) 114–131.

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

[23] R.B. D'Agostino, R.S. Vasan, M.J. Pencina, P.A. Wolf, M. Cobain, J.M. Massaro, W.B. Kannel, General cardiovascular risk profile for use in primary care. Circulation 117 (6) (2008) 743–753, https://doi.org/10.1161/circulationaha.107.699579.

[24] P. Chatterjee, L. Cymberknop, R. Armentano, IoT-based ehealth toward decision support system for CBRNE events. in: A. Malizia, M. D'Arienzo (Eds.), Enhancing CBRNE Safety & Security: Proceedings of the SICC 2017 Conference, Springer, Cham, 2018. https://doi.org/10.1007/978-3-319-91791-7_21.

[25] Understanding MELD Score for Liver Transplant: UPMC. (n.d.). Retrieved March 21, 2020, from https://www.upmc.com/services/transplant/liver/process/waiting-list/meld-score.

[26] P. Chatterjee, L.J. Cymberknop, R.L. Armentano, 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, https://doi.org/10.1109/CSNT.2017.8418567.

[27] P. Chatterjee, R.L. Armentano, L.J. Cymberknop, 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, https://doi.org/10.1109/MLDS.2017.22.