MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231094

Characterizing Cluster-Based Frailty Phenotypes in a Multicenter Prospective Cohort of Kidney Transplant Candidates

Syed Hani Raza ABIDI[a], Nur ZINCIR-HEYWOOD[a], Syed Sibte Raza ABIDI[a], Kranthi JALAKAM[a], Samina ABIDI[c], Lakshman GUNARATNAM[d], Rita SURI[e], Héloïse CARDINALE[f], Amanda VINSON[b], Bhanu PRASAD[g], Michael WALSH[h], Seychelle YOHANNA[h], George WORTHEN[b] and Karthik TENNANKORE[b,1]
[a]*Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada*
[b]*Division of Nephrology, Dept. of Medicine, Dalhousie University, Halifax, Canada*
[c]*Dept. of Community Health & Epidemiology, Dalhousie University, Halifax, Canada*
[d]*Division of Nephrology, London Health Sciences Center, London, Ontario, Canada*
[e]*Faculty of Medicine, McGill University, Montreal, Quebec, Canada*
[f]*Division of Nephrology, Centre de Recherche du CHUM, Montreal, Quebec, Canada*
[g]*Division of Nephrology, Regina General Hospital, Regina, Saskatchewan, Canada*
[h]*Division of Nephrology, McMaster University, Hamilton, Ontario, Canada*

Abstract. Frailty is associated with a higher risk of death among kidney transplant candidates. Currently available frailty indices are often based on clinical impression, physical exam or an accumulation of deficits across domains of health. In this paper we investigate a clustering based approach that partitions the data based on similarities between individuals to generate phenotypes of kidney transplant candidates. We analyzed a multicenter cohort that included several features typically used to determine an individual's level of frailty. We present a clustering based phenotyping approach, where we investigated two clustering approaches—i.e. neural network based Self-Organizing Maps (SOM) with hierarchical clustering, and KAMILA (KAy-means for MIxed LArge data sets). Our clustering results partition the individuals across 3 distinct clusters. Clusters were used to generate and study feature-level phenotypes of each group.

Keywords. Frailty, kidney transplant, phenotypes, machine learning, clustering

1. Introduction

Frailty has been shown to be associated with a higher risk of death among dialysis patients, kidney transplant recipients and even those who are waiting for a kidney

[1]Corresponding Author: Dr. Karthik Tennankore. Email: email: KarthikK.Tennankore@nshealth.ca.

transplant [1, 2]. Among waitlist candidates, the measurement of frailty often involves established frailty measurement tools (i.e. the Fried Phenotype), clinical impression or a comprehensive assessment based on a summation of deficits across several domains (including social, function, mobility, health and cognition), known as a frailty index [3, 4]. These tools all have advantages but may be informed by only a limited number of variables[5]. Partitioning the data based on similarities between individuals, when informed by a wide range of variables, may better help to identify waitlist candidates who are at higher risk by virtue of having a high proportion of deficits.

In this paper, we present a clustering based phenotyping approach, where we investigated two different clustering approaches applied to a multicenter, prospective dataset of kidney transplant candidates with multi-type data and a large number of features. We investigated neural network based self-Organizing Maps (SOM) [6] with hierarchical clustering, and KAMILA (KAy-means for MIxed LArge data sets), a density based clustering method [7]. Our clustering results show that the data has 3 distinct clusters which were further analyzed using statistical methods to understand feature-level phenotypes of waitlist candidates grouped in each cluster.

2. Methods

2.1. Data description

We analyzed a cohort of 807 kidney transplant recipients who were previously recruited to a multicenter prospective study evaluating the impact of frailty on outcomes for kidney transplant waitlist candidates (from 2016-2022) [8]. The dataset only included those individuals who were within 6 months of their date of being placed on the kidney transplant waitlist (N=367) for whom there were 105 variables across a number of domains. Patients in the original study were recruited from five Canadian transplant centers located in London (Ontario), Hamilton (Ontario), Halifax (Nova Scotia), Saint John (New Brunswick), and Montreal (Quebec). Variables in the dataset included demographics, comorbid health conditions, self-reported functional limitation, both self-reported and measured variables relating to cognitive function, social and emotional well-being, and physical examination of low grip strength and slow walking time. The dataset contained both categorical and continuous features.

2.2. Data Clustering

We used ML based clustering methods to partition the data into distinct clusters, where each cluster represents patients with similar characteristics. We investigated neural network based Self-Organizing Maps (SOM) as they have the inherent properties of dimension reduction to a 2-diemnsional map through vector quantization, and visualization of the clusters. For SOM we used the *aweSOM* library in R studio. We applied the iterative clustering method KAMILA that uses kernel density

estimation for continuous features and multinomial modelling for categorical features. Clustering performance, was evaluated using the silhouette score, Dunn Index, and Calinski-Harabasz (CH) Index. Silhouette score determined the optimal number of clusters. Best clustering result was determined based on both silhouette and CH index.

3. Results

Table 1 presents the results for SOM based clustering experiments. The ideal SOM had a dimension of 7x7, and the data was distributed into 3 reasonably sized clusters. Figure 1 illustrates the clustering result generated by the 7x7 SOM. The experimental results using KAMILA, showed sub-optimal cluster distributions with one dominant and two small clusters. Therefore, the SOM based clusters were used for generating phenotypes for the cohort of waitlist candidates.

Table 1: Results of SOM based clustering experiments.

Dimensions	Linkage	Cluster Sizes	Silhouette	Dunn	CH
7x7	**Complete**	**198, 90, 79**	**0.1124315**	**0.1273989**	**39.7**
10x10	Ward D.2	188, 131, 48	0.0878228	0.1323267	42.9
12x12	Ward D.2	202, 51, 114	0.0911939	0.07530532	41.9
15x15	Ward D	203, 50, 114	0.1000332	0.1334696	40.5
17x17	Ward D.2	87, 67, 213	0.115904	0.1387107	39.6

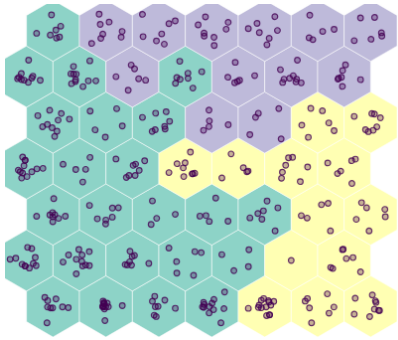

Figure 1. Distribution and structure of the three clusters generated by SOM.

4. Discussion

We analyzed the distribution of characteristics within each cluster to generate phenotypes across three main dimensions: Comorbidity, Functional/Mobility impairment, and Mental and Social Function. Table 2 illustrates the phenotype for each cluster based on the values of select features. The cells highlighted in red

indicate a known high risk feature, orange cells indicate medium risk, and green cells indicate low-risk. According to the analysis, cluster 3 (79 patients) represents patients with the highest number of impairments across all domains, followed by cluster 2 (N=90 patients), and finally cluster 1 (N=198 patients).

Table 2. Baseline Characteristics of Patients Stratified by Cluster Assignment.

Variable	Cluster 1	Cluster 2	Cluster 3	P
	N=198	N=90	N=79	
Demographics				
Age (years +/- standard deviation)	55 +/-13	48 +/-13	59 +/-12	**<0.001**
Comorbidity: N (%)				
Diabetes	68 (34)	8 (9)	54 (68)	**<0.001**
Coronary Artery Disease	33 (17)	3 (3)	26 (33)	**<0.001**
Cerebrovascular Disease	8 (4)	8 (9)	11 (14)	**0.014**
Prior Cancer	21 (11)	2 (2)	13 (16)	**0.007**
Functional : N (%)				
Exhaustion	28(14)	44 (49)	49 (62)	**<0.001**
Weak Grip	58 (29)	28 (31)	41 (52)	**0.001**
Slow Walking Speed	15 (8)	15 (17)	18 (24)	**0.001**
Trouble lifting 10 pounds	3 (2)	11 (12)	14 (17)	**<0.001**
Help with Housework	11 (6)	24 (27)	23 (29)	**<0.001**
Help with Groceries	19 (10)	38 (42)	40 (51)	**<0.001**
Help Climbing Stairs	0	1 (1)	11 (14)	**<0.001**
Limited in Walking 100 Metres	6 (3)	5 (6)	22 (28)	**<0.001**
Limited in Moderate Activity	25 (13)	46 (51)	64 (81)	**<0.001**
Limited in Climbing One Flight of Stairs	6 (3)	4 (4)	36 (46)	**<0.001**
Cuts Work Due to Physical Function	71 (36)	64 (71)	61 (77)	**<0.001**
Social, Emotional, Cognitive				
Feels Everything is an Effort	37 (19)	73 (81)	56 (71)	**<0.001**
Feels Alone	16 (8)	32 (36)	30 (38)	**<0.001**
Cannot recall 3 words	70 (36)	16 (18)	25 (32)	**0.010**
Rarely Socializes	60 (30)	45 (50)	49 (62)	**<0.001**
Cuts Work Due to Emotional Reasons	2 (1)	9 (10)	14 (18)	**<0.001**

Analysis of Cluster 1: The largest group of patients (198 patients) have the fewest issues with functional limitation. From a social and emotional perspective they tended to have the least degree of impairment, with the exception of cognitive recall; a sizable portion of them had trouble recalling 3 words. In terms of comorbidity, these patients were at intermediate risk with high proportions of prior coronary artery disease (17%) and diabetes (34%). Overall, patients in this cluster were mostly independent and emotionally healthy, but with a moderate burden of comorbidity.

Analysis of Cluster 2: This group of 90 patients was the youngest and contained the highest proportion of individuals of female sex (60%). Patients in this cluster had a minimal burden of comorbidity. In terms of functional status patients in this cluster did require a degree of assistance for a number of tasks, a tendency to exhaustion and impaired grip strength. Interestingly, these patients did exhibit impaired social, and emotional function, most notably in perceived effort, although cognitive recall was

not as impaired as for those in clusters 1 and 3. Patients in this cluster have enough functional capacity to perform many basic tasks, they have very few comorbidities emphasizing that functional impairment is not synonymous with medical conditions.

Analysis of Cluster 3: The smallest cluster (79 patients) is the highest-risk cluster made up of the largest percentage of patients with severe health problems such as diabetes (68%), coronary artery disease (33%), and prior malignancy (16%). In terms of functional status, this group had trouble performing basic tasks such as walking, bathing, climbing stair and doing groceries. With respect to emotional and social function, this cluster has the largest percentage of patients who felt lonely and socially isolated. Overall, this group was characterized by individuals with a high frailty severity and impairments across a number of health domains.

5. Conclusions

ML-based clustering provides an automated approach to generate phenotypes of kidney transplant candidates using a large number of multi-type variables. Our clustering methods grouped similar patients, given a heterogeneous population, to generate phenotypes based on all the patient features, as opposed to just some important features which is the case with manual phenotyping. The generated phenotypes can be used to identify waitlist candidates who are at higher frailty risk. Frailty is important for all organ transplant candidates and recipients, as such our approach can be applied broadly to other organ transplants.

References

[1] Alfaadhel TA, Soroka SD, Kiberd BA, Landry D, Moorhouse P, Tennankore KK. Frailty and mortality in dialysis: evaluation of a clinical frailty scale. Clin J Am Soc Nephrol. 2015 May;10(5):832-40, doi: 10.2215/CJN.07760814.

[2] McAdams-DeMarco MA, Law A, King E, Orandi B, Salter M, Gupta N, Chow E, Alachkar N, Desai N, Varadhan R, Walston J, Segev DL. Frailty and mortality in kidney transplant recipients. Am J Transplant. 2015 Jan;15(1):149-54, doi: 10.1111/ajt.12992.

[3] Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Seeman T, Tracy R, Kop WJ, Burke G, McBurnie MA; Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001 Mar;56(3):M146-56.

[4] Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I, Mitnitski A. A global clinical measure of fitness and frailty in elderly people. CMAJ. 2005 Aug;173(5):489-95.

[5] Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatr. 2008 Sep;8:24, doi: 10.1186/1471-2318-8-24.

[6] Kohonen, Teuvo. Self-organizing maps. Vol. 30. Springer Science & Business Media, 2012.

[7] Foss, A., Markatou, M., Ray, B. et al. A semiparametric method for clustering mixed data. Mach Learn. 2016;105: 419–58, doi.org/10.1007/s10994-016-5575-7.

[8] Tennankore KK, Gunaratnam L, Suri RS, Yohanna S, Walsh M, Tangri N, Prasad B, Gogan N, Rockwood K, Doucette S, Sills L, Kiberd B, Keough-Ryan T, West K, Vinson A. Frailty and the Kidney Transplant Wait List: Protocol for a Multicenter Prospective Study. Can J Kidney Health Dis. 2020 Sep;7:2054358120957430, doi: 10.1177/2054358120957430.