MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190230

Visual Analytics for Congestive Heart Failure Mortality Prediction

Rema Padman[a], Ofir Ben-Assuli [b], Tsipi Heart[b], Nir Shlomo [c], Robert Klempfner[c]

[a] The H. John Heinz III College of Information Systems and Public Policy, Carnegie Mellon University, Pittsburgh, USA
[b] Faculty of Business Administration, Ono Academic College, Kiryat Ono, Israel
[c] Heart Center, Sheba Medical Center, Ramat Gan, Israel

Abstract

Several indices exist to classify Congestive Heart Failure (CHF) patients' propensity for early mortality; however, they are primarily based on limited data and are not intuitive to use at the point of care. We investigate a novel, data-driven, risk assessment and visualization approach to investigate mortality prediction of CHF patients using data retrieved from an intensively digitized hospital's data repository. Combining well-known, computationally efficient, dimensionality reduction (DR) methods with 2-d information visualization, the method classifies and visualizes CHF patients into high and low risk groups, contextualized by the factors driving their classification. The DR method performed similar to logistic regression (LR), but visualized the classification and its significant factors at the population level, individual level and the potential impact of interventions for an individual patient. These are encouraging results in favor of the proposed visualization approach, and contributes to the current focus on advancing patient care via large-scale visual analytics.

Keywords:

risk assessment, computer graphics, heart failure

Introduction

The widespread adoption of Health Information Technology (HIT) by healthcare organizations in many parts of the world has generated vast repositories of detailed patient-level data as a result of routine care delivery [1]. Efficient and effective use of multi-dimensional, multi-sourced data for clinical decision support at the point of care is further exacerbated by the time-constrained and stressed environments where most of health care is provided – in the general practitioner (GP)'s office, hospitals and emergency care institutions [2]. Visually intuitive methods and tools that support clinicians for early identification of high-risk patients with Congestive Heart Failure (CHF), as well as patients with better adherence to treatment recommendations and shared decision-making, can improve both clinical decision-making and patient health outcomes. The objective of this study is to demonstrate a multi-level visualization of CHF patients' disease risk classification that combines statistical dimensionality reduction (DR) methods with innovative information visualization to enhance efficient and effective clinical decision support at the point of care. The visual presentation provides a simple, intuitive tool for clinicians to quickly identify high-risk patients and understand key factors affecting their condition at a snapshot in time.

Congestive Heart Failure (CHF)

Based on recent data, nearly 5 million US adults are currently living with CHF as are more than 25 million adults worldwide [3]. CHF affects people of all ages, from children and young adults to the middle-aged and elderly. Almost 1.4 million individuals with CHF in the U.S are under 60 years of age. CHF is responsible for 11 million physician visits each year in the US, and more hospitalizations than all forms of cancer combined (https://www.emoryhealthcare.org/heart-vascular/wellness/heart-failure-statistics.html). In fact, CHF is the leading cause of death for both men and women, and the most common diagnosis in hospital patients age 65 years and older. More than half of those who develop CHF die within 5 years of diagnosis, and sudden death is common in patients with CHF, occurring at a rate of six to nine times that of the general population [4, https://www.cdc.gov/heartdisease/facts.htm].

Several indices exist to classify CHF patients' propensity for early mortality. Among the most popular are the Enhanced Feedback for Effective Cardiac Treatment (EFFECT) [5]; the Emergency department-based models of Society of Chest Pain Centers, and the Ottawa Heart Failure Risk Scale [6].

Current prediction tools are largely based on limited data, rather than on patient-level big data derived from various data sources [7, 8]. Furthermore, most prior studies either used explanatory rather than predictive statistics, or utilized traditional, albeit sometimes complex, statistical models [9, 10].

Visual Analytics (VA)

Data or information visualization refers to graphical *presentation* of complex computerized results in formats that enable rapid comprehension of complex situations by the viewers, but does not necessarily involve *analysis* tasks or *interactivity* [11]. In contrast, VA requires collaborative interaction between the user and the computer. It is defined as the science of analytical reasoning facilitated by interactive visual interfaces [12]. VA combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision-making on the basis of very large and complex data sets. Using VA, the results are presented via visual interactive platforms for decision makers to explore and gain insights and new knowledge. The approach is to marry the big data processing capabilities of analytics with the human intuitive capabilities of interactive visualization.

Benefits of using visualization tools

Innovative algorithms and tools that combine statistical machine learning and optimization with information visualization and electronic health data may reduce clinicians' information processing load and improve their ability to assess risk of disease onset and related complications at the point of care [2, 13]. Information visualization utilizes the high bandwidth processing capabilities of the human visual system to more efficiently perform interactive data exploration and glean important insights [14, 15]. A critical element in visualization is the incorporation of an expert user, such as the clinician decision-maker, in the interpretation of the data. This may make visualization methods particularly useful for cardiology care where clinicians desire flexibility for customizing assessments to the needs of their unique patient

populations. To the best of our knowledge, this is the first study on computationally driven visualization techniques for improving CHF risk assessment.

We applied the visual analytics method to two subsets of the uniquely comprehensive cardiology dataset from a major academic medical center in Israel. We compare the risk visualization approach with Logistic Regression (LR) results to highlight the unique value of visual analytics in clinical decision making for complex health conditions.

Methods

Study setting and data

The Sheba Medical Center is the largest Israeli hospital, located in the center of the country. It is a leading public academic medical center, annually handling over a 1.5 million patient visits and ~200K emergency visits, and conducts more than two million medical tests of all types. Sheba is an extensively computerized hospital, where an integrated EMR system is used in the ED and all inpatient departments, fully replacing paper-based medical records. In recent years, Sheba has established a research-focused data warehouse which collects data describing imaging results, manual monitored parameters keyed into the clinical information system in the intensive care unit (patient temperature, blood pressure, pulse, blood oxygen saturation, weight, height, etc.), medical devices data (ECG, Echocardiography examinations, Cardiac Catheterization, Nuclear Imaging), the National Population Registry, the National Cancer Registry and the Sheba Executive Survey data.

Each patient admitted to the hospital, via the ED or otherwise, who is diagnosed with a cardiovascular condition as a primary diagnosis, has data uploaded to the data warehouse. All patient information is then retrieved from the other systems using the unique patient ID as it appears in the Israeli Population Registry. This unique identifier allows accurate location of patient data. The records have the potential to contain above thousand features for each patient, albeit sparsely populated. Such a health data repository is both rich and unique for large-scale data analysis, but there are many data challenges. While return of investment of such data integration projects is uncertain [16], one of the use cases proposed for health data is to improve clinicians' decisions at the point of care.

In this preliminary study, we use a subset of the uniquely comprehensive Sheba data warehouse for our analysis, which provides many relevant variables affecting readmissions and mortality and covers all possible patients treated at Sheba over the recent five years. Consequently, this represents the whole population of interest instead of a limited, skewed sample.

Data sample

We employed k-means cluster analysis on a dataset of 7,168 patients hospitalized at the Sheba Medical Center for heart disease between 2010 and 2017. For the demonstration of the visual analytics approach, we selected two specific clusters: younger (N= 204, average age 63 (±15.5)) and older patients (N=367, average age 79.12 (±10.57). Table 1 shows the descriptive statistics of the younger cohort. The other cohort is omitted due to page limit constraints.

Logistic regression to predict patients' propensity for early mortality

Among the many analytical methods currently available, we chose logistic regression (LR) as a baseline method to compare with the visual analytics approach. LR is a widely used and accepted statistical method in clinical medicine [17].

Visual analytics approach

We apply prior work on dimensionality reduction and information visualization for disease risk assessment [2, 13] to CHF data. Specifically, the visual analytics approach incorporates Principal Component Analysis (PCA) and Fisher's Linear Discriminant Analysis (LDA) to develop informative two-dimensional (2-D) projections of multi-dimensional patient data and classify them for CHF-related risk assessment [2, 13]. Included in this step are the identification of appropriate data normalization procedures for the dimensionality reduction methods and the disparate measurement of the data attributes. A set of feasible methods for pre-processing and projecting high-dimensional patient data to 2-D plots are incorporated into the framework so that multiple visual enhancements that may augment a user's analysis can be obtained. Prior results show that the framework may generate models which visually classify a large patient population with accuracy comparable to that produced by common statistical methods [2, 13].

Starting with a d-dimensional space (the d features of interest in the current study), PCA uses an eigenvalue decomposition to find d orthogonal linear combinations that explain the most variance in the data [18]. The principal components are ordered based on the amount of variance explained. LDA, on the other hand, is explicitly concerned with classification and takes into account data labels [19]. In a dataset labeled with k classification levels, LDA finds a k-1 dimensional projection of the d-dimensional data that maximizes the ratio of between-class variance to within-class variance. LDA was chosen because it stratifies patients into risk groups. PCA ensures that patient observations are maximally scattered in the reduced space. The data are projected into 2-D space as scatter plots using the optimal linear combinations found by PCA and LDA, with LDA finding a decision boundary for classifying patients into high and low risk categories. Contextualization of the scatter plot is critical for ease of use and interpretability, and is provided by anchoring the risk factors around the scatter plot, effectively circumscribing them in a circle. Each anchor's direction of attraction (θi) is defined using the ratio of that feature's weights on the vertical and horizontal components of the two-dimensional space:

$$\theta_i = \begin{cases} \tan^{-1}(d_{i1} / d_{i2}) & d_{i2} \geq 0 \\ \pi + \tan^{-1}(d_{i1} / d_{i2}) & d_{i2} < 0 \end{cases} \quad (1)$$

where d_{i1} is the weight of feature i on the vertical axis component, and d_{i2} is the weight of feature i on the horizontal axis component. Finally, the significance of each anchor can be represented by the size of the anchor (S_i), computed, for simplicity, as proportional to either d_{i1} or d_{i2}. Further details of the method and its evaluation with diabetes data are reported in [2, 13].

These methods have been implemented in an RShiny-based platform (https://shiny.rstudio.com/) that allows loading, viewing, descriptive analyses, predictive analyses and visual analytics of any dataset. We employ these methods and tools to a subset of the Sheba data sample with a set of selected features to investigate the population distribution between low- and high-risk patients, and examine factors emerging as most critical for high-risk patients. The risk prediction results from this phase are compared against the results from LR to generate additional insights, particularly to identify patients who are noted to be at the boundary of high and low risk levels by the two approaches.

Results

The two data subsets represent older group (OG) and younger group (YG) of CHF patients, on average, that are randomly retrieved from the large dataset. In the future, we intend to use the complete dataset, possibly even over a longer duration, and include additional data on demographics, diagnoses and current and past health indicators, medications, and lab and imaging results.

Descriptive statistics

The mean and proportion of indicators that differed significantly between the younger group (YG- alive and deceased) are highlighted in Table 1 (The older group (OG) is also discussed but the table is not included due to page limits). As evident in Table 1, among factors affecting significantly likelihood of mortality are age (older more likely), prior CHF, dyslipidemia, hypertension and diabetes mellitus. Among lab tests, patients with lower level of hemoglobin in the first test (indication of anemia), and lower GFR MDRD (measure of renal function) were more likely to die. Additional factors were lower systolic and diastolic blood pressure among patients who died. Among the older patients (not shown as a table), age was a significant factor (older more likely to die), lower GFR MDRD and lower systolic and diastolic blood pressure. For this cohort, however, two unique factors affected the likelihood of older patients' mortality: higher sodium level in the blood and lower level of hematocrit. Interestingly, factors significantly affecting the likelihood of mortality in younger patients were derived from medical history and risk factors, whereas for older patients they were derived mainly from lab tests and physical examination.

Table 1: Data description of the younger population

Indicator and Description	Total	Alive (0)	Deceased (1)	P value
N:Total admissions included	204	160	44	
Patient age (mean ± sd)	63 (15.6)	61 (15.5)	70 (13.4)	<0.001
Male (%)	144 (71)	114 (71)	30 (68)	0.835
Ejection fraction (mean ± sd)	38.08 (17)	37.38 (17)	40.66 (16)	0.266
Past Percutaneous Coronary Infusion (%)	48 (23)	36 (22)	12 (27)	0.645
Past Myocardial Infarction (%)	55 (27)	44 (27)	11 (25)	0.889
Congestive Heart Failure (%)	117 (57.4)	90 (56.2)	27 (61.4)	0.663
Past Stroke (%)	15 (7.4)	13 (8.1)	2 (4.5)	0.632
Peripheral arterial disease (%)	7 (3.4)	5 (3.1)	2 (4.5)	1.000
Dyslipidemia (%)	82 (40.2)	58 (36.2)	24 (54.5)	0.044
Hypertension (%)	101 (49.5)	73 (45.6)	28 (63.6)	0.052
Diabetes Mellitus (%)	64 (31.4)	43 (26.9)	21 (47.7)	0.014
Chronic Obstructive Pulmonary Disease (%)	10 (4.9)	6 (3.8)	4 (9.1)	0.290
FIRST Hemoglobin (mean ± sd)	12.49 (1.91)	12.76 (1.81)	11.52 (1.96)	<0.001
Left Ventricular Function (%)	133 (65.2)	107 (66.9)	26 (59.1)	0.435
SPAP (mean ± sd)	47.33 (13.00)	47.09 (13.46)	48.23 (11.25)	0.606
Glomerular Filtration Rate MDRD (mean ± sd)	60.39 (30.35)	67.40 (29.10)	34.91 (19.24)	<0.001
SODIUM (mean ± sd)	138.01 (4.23)	137.91 (3.59)	138.39 (6.05)	0.506
Hematocrit (mean ± sd)	37.04 (6.33)	37.20 (5.65)	36.43 (8.39)	0.478
ACE/ARBs (%)	176 (86.3)	141 (88.1)	35 (79.5)	0.223
Chronic Aldactone (%)	127 (62.3)	105 (65.6)	22 (50.0)	0.086
Length of Stay (mean ± sd)	8.60 (7.11)	8.27 (6.72)	9.81 (8.36)	0.206
Atrial Fibrillation (%)	52 (25.5)	37 (23.1)	15 (34.1)	0.200
Systolic Blood Pressure (mean ± sd)	120.27 (24.00)	124.82 (21.44)	103.73 (25.74)	<0.001
Diastolic Blood Pressure (mean ± sd)	67.82 (14.21)	70.86 (11.86)	56.80 (16.56)	<0.001
First Glucose (mean ± sd)	123.00 (54.19)	120.15 (52.06)	133.36 (60.83)	0.153
First SGOT (mean ± sd)	37.78 (45.01)	39.48 (49.53)	31.61 (21.03)	0.306
First INR (mean ± sd)	1.45 (0.81)	1.42 (0.80)	1.54 (0.81)	0.385
Echo Fractional Shortening (mean ± sd)	26.05 (12.33)	25.28 (12.22)	28.85 (12.49)	0.089
Body Mass Index (mean ± sd)	27.73 (4.51)	27.73 (4.65)	27.70 (3.98)	0.966
Echo LV Mass Index (mean ± sd)	117.14 (29.05)	118.21 (29.72)	113.24 (26.38)	0.316

Table 2: Comparing DR and LR on statistically significant factors driving risk

DR		LR	
Mortality Younger	Mortality Older	Mortality Younger	Mortality Older
Pre-MI	Any-RE-CHF	DM	HCT
Pre-CVA	First-HGB	GFR MDRD (lower)	GFR MDRD (lower)
Pre-COPD	SPAP, Sodium level	Sodium level	Sodium level
First HGB	DBP, Glucose	SBP (lower)	SBP (lower)
HCT	Any-ARBs		
Any-ACE	HOS-Duration		
Pre-A.Fib	Pre-AFib		

Table 3: Comparison across Dimensionality Reduction (DR) vs. Logistic Regression (LR)

Mortality: Young vs. Old Patients - Comparison across methods				
	Young		Old	
Performance measures	DR	LR	DR	LR
Accuracy	0.853	0.863	0.668	0.807
Recall	0.864	0.795	0.867	0.722
Precision	0.613	0.648	0.415	0.586
F1 Score	0.717	0.714	0.561	0.647
Confusion matrix				
Actual / Predicted	DR	LR	DR	LR
Died / High risk patients (TP)	20	35	78	65
Died / Low risk patients (FN)	0	9	12	25
Not died / High risk patients (FP)	60	19	110	46
Not died / Low risk patients (TN)	124	141	167	231

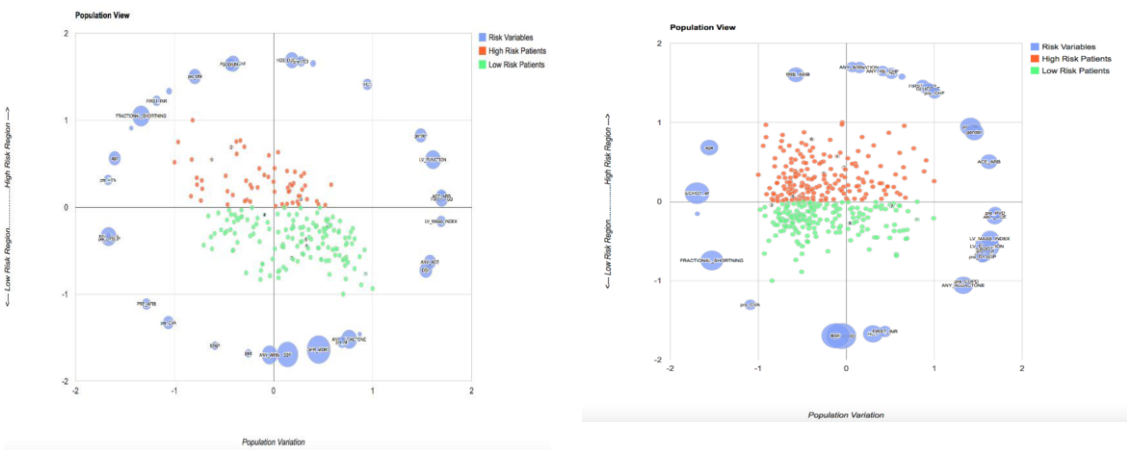

Figure 1: Population-level mortality risk prediction for younger (top) and older (bottom) patient clusters

Dimensionality reduction and population risk visualization

Figure 1 displays the population-level mortality risk prediction for younger and older patients, respectively, using the statistical dimensionality reduction approach described above. It is interesting to note that the distribution of mortality risk in the respective patient cohorts (indicated by red points for high risk and green for low risk patients), the factors driving risk in each group (indicated by the blue circles at the top of the displays and listed in Table 2) and protective factors (indicated by the blue circles at the bottom of the displays – compliance with medications) are distinct and different for each subset. These visualizations provide new insights about customized, cohort-driven population risk management. Individual patient-level and intervention-level analysis can be performed analogously, as described in [2, 13].

The DR technique classified almost 70% of the younger group to have low-risk of mortality, while 30% were classified as high-risk (Figure 1, top). In this cohort, there are seven factors in the high risk region (top of the graph) that are driving risk for this population while eight factors at the bottom of the graph are protective.

Comparison of DR and LR

Table 2 shows factors that have been determined to be statistically significant in driving the classification as high or low risk. It is interesting to note that that DR identifies a different set of factors for the younger vs. the older populations whereas the factors are quite similar in the LR solution. DR also identifies a larger group of modifiable factors that clinicians can intervene on compared to LR. These findings need to be evaluated with larger data sets and validated extensively with practicing clinicians and the clinical literature.

We compared the two methods using three measures (Table 3): Accuracy (TP + TN divided by the number of participants), Recall (TP divided by TP + FN) and Precision (TP divided by TP + FP). DR outperformed LR on the Recall measure for both cohorts, an important indicator for mortality.

Discussion

The DR analysis provides a visual classification of the patient cohort into low- and high-risk regions. Factors affecting the specific classification can be derived and visualized as anchors around the scatter plot (circles), as can the magnitude of the effect of individual anchors (size of the circle). It is important to demonstrate the validity of this classification method in the context of multiple disease conditions. This has been done by comparing the visual analytic results to those of logistic regression, a common classification method. As presented in Tables 1, 2 and 3, DR outperformed LR in discriminating high risk patients (TP), while LR performed better in identifying low risk patients (TN). Although results of other comparison measures (accuracy, precision) were inconclusive, the capacity of DR to more precisely identify high-risk patients, without likewise increasing the falsely identified patients (especially FN) is encouraging (Table 3). Minimizing the FN is the major criterion in actual practice. Moreover, DR identified a higher number of significant factors affecting high-risk patients specifically for older, readmitted patients (Table 2).

While expert physicians are generally able to classify regular patients to low- or high-risk categories, patients with multiple comorbidities often are more difficult to classify correctly. We found that the DR tool was able to classify patients as either low- or high-risk, thus aiding the clinician to refine their assessment. Both options are important: classifying low-risk patients as high-risk is costly and causes unnecessary worry and inconvenience for the patients, and classifying high-risk patients as low-risk may be dangerously life threatening. This approach offers a unique opportunity to deliver cognitively-guided capabilities that have the potential to move statistical risk models closer to the frontlines of clinical practice using visualization techniques, and facilitate the goal of meeting clinician and patient information needs to improve care quality and health outcomes. The demonstrated visualization techniques may benefit multiple stakeholders, including patients, clinical practitioners, researchers and policy-makers, with generalizability to many risk assessments in clinical practice.

Conclusions

We have demonstrated the promise of an innovative statistical model-based approach to create a visualization tool (dimensionality reduction and 2-D visualization). The advantages are evident in the lower number of FNs and identification of modifiable significant factors that differ between younger and older populations. In terms of model characteristics, such as accuracy and recall, results are not conclusive. Additional analyses with larger patient cohorts need to be carried out.

Future research includes many avenues such as extending the methodology to allow time-varying risk factors, optimizing feature selection and decision boundary alignment and evaluation against several other statistical and machine learning methods.

References

[1] E. Scheurwegs, K. Luyckx, L. Luyten, W. Daelemans, and T. Van den Bulcke, Data integration of structured and unstructured sources for assigning clinical codes to patient stays, *JAMIA* 2016, doi: 10.1093/jamia/ocv115.

[2] C.A. Harle, D.B. Neill, and R. Padman, Information visualization for chronic disease risk assessment, *IEEE Intelligent Systems* **27** (2012), 81-85.

[3] G. Savarese and L.H. Lund, Global Public Health Burden of Heart Failure, *Cardiac failure review* **3** (2017), 7-11.

[4] American Heart Association, Cardiovascular Disease: A Costly Burden for America, Projections Through 2035, in, www.healthmetrics.heart.org, 2017.

[5] F.J. Martín-Sánchez, V. Gil, P. Llorens, P. Herrero, J. Jacob, C. Fernández, Ò. Miró, and A.H.F.W.G.o.t.S.S.o. E.M.I. Group, Barthel Index—Enhanced Feedback for Effective Cardiac Treatment (BI—EFFECT) Study: Contribution of the Barthel Index to the Heart Failure Risk Scoring System Model in Elderly Adults with Acute Heart Failure in the Emergency Department, *Journal of the American Geriatrics Society* **60** (2012), 493-498.

[6] I.G. Stiell, J.J. Perry, C.M. Clement, R.J. Brison, B.H. Rowe, S.D. Aaron, A.D. McRae, B. Borgundvaag, L.A. Calder, and A.J. Forster, Prospective and Explicit Clinical Validation of the Ottawa Heart Failure Risk Scale, With and Without Use of Quantitative NT-pro BNP, *Academic Emergency Medicine* **24** (2017), 316-327.

[7] S.E. Mitchell, J. Martin, S. Holmes, C.van D. Lukas, R. Cancino, M. Paasche-Orlow, C. Brach, and B. Jack, How Hospitals Reengineer Their Discharge Processes to Reduce Readmissions, *J Health Qual*. **38** (2016), 116-126.

[8] A.J. Sommerfeld, A.D. Althouse, J. Prince, and G.W. Hickey, Obstructive Sleep Apnea is Associated with Increased Readmissions in CHF Patients, *Journal of Cardiac Failure* **22** (2016), S88.

[9] R. Amarasingham, F. Velasco, B. Xie, C. Clark, Y. Ma, S. Zhang, D. Bhat, B. Lucena, M. Huesch, and E.A. Halm, Electronic medical record-based multicondition models to predict the risk of 30 day readmission or death among adult medicine patients: validation and comparison to existing models, *BMC Medical Informatics and Decision Making* **15** (2015), 39.

[10] J. Futoma, J. Morris, and J. Lucas, A comparison of models for predicting early hospital readmissions, *Journal of Biomedical Informatics* **56** (2015), 229-238.

[11] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, Visual analytics: Definition, process, and challenges, In: Kerren A., Stasko J.T., Fekete J.D., and North C., (eds.), *Information visualization, Lecture Notes in Computer Science, vol 4950.* Springer, Berlin, Heidelberg, 2008, pp. 154-175.

[12] K.A. Cook and J.J. Thomas, Illuminating the path: The research and development agenda for visual analytics, in, Pacific Northwest National Lab., Richland, WA, 2005.

[13] C.A. Harle, D.B. Neill, and R. Padman, An information visualization approach to classification and assessment of diabetes risk in primary care, in: *Proceedings of the 3rd INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2008)*, 2008.

[14] J.J. Caban and D. Gotz, Visual analytics in healthcare–opportunities and research challenges, Oxford University Press, 2015.

[15] B. Shneiderman, The eyes have it: A task by data type taxonomy for information visualizations, In: *Proceedings, IEEE Symposium on Visual Languages*, 1996, pp. 336-343.

[16] L.E. Price, K. Shea, and S. Gephart, The Veterans Affairs's Corporate Data Warehouse: Uses and Implications for Nursing Research and Practice, *Nursing Administration Quarterly* **39** (2015), 311-318.

[17] A. Dag, A. Oztekin, A. Yucel, S. Bulur, and F.M. Megahed, Predicting heart transplantation outcomes through data analytics, *Decision Support Systems* **94** (2017), 42-52

[18] I. Joliffe, Principal Component Analysis. Springer, 2002.

[19] T. Hastie, J. Friedman, and R. Tibshirani, Linear methods for classification, *The Elements of Statistical Learning*, Springer, 2001, pp. 79-113.

Address for correspondence

Prof. Rema Padman, The H. John Heinz III College of Information Systems and Public Policy, Carnegie Mellon University, Pittsburgh, USA, rpadman@cmu.edu.