KNOWLEDGE
MEDIA
DESIGN
INSTITUTE

KMDI REPORTS

Predicting ICU Death with Summarized Data: The Emerging Health Data Search Engine

Authors:
Mahsa Rouzbahman, PhD Candidate,
and Mark Chignell, Professor,

KMD-2014-4

Predicting ICU Death with Summarized Data: The Emerging Health Data Search Engine

Mahsa Rouzbahman, *PhD Candidate,* and Mark Chignell, *Professor, University of Toronto*

***Abstract*— Chignell et al. [1] previously described a methodology for converting a large set of confidential data records into a set of summaries of similar patients. They claimed that the resulting patient types could "capture important trends and patterns in the data set without disclosing the information in any of the individual data records." In this paper we examine the predictive validity of an initial set of patient types developed by [1]. We ask the following question: To what extent can the summarized data derived from each cluster (patient type) be as informative as the original case level data (individuals) from which the clusters were inferred? We address this question by assessing how well predictions made with summarized data matched predictions made with original data. After reviewing relevant literature, and explaining how data is summarized in each cluster of similar patients, we compare the results of predicting death in the ICU[1] using both summarized (regression analysis) and original case data (discriminant analysis and logistic regression analysis). When multiple clusters were used, prediction based on regression analysis of the summarized data was found to be better than prediction using either logistic regression or discriminant analysis on the raw data. We hypothesize that this result is due to segmentation of a heterogenous multivariate space into more homogeneous subregions. We see the present results as an important step towards the development of generalized health data search engines that can utilize non-confidential summarized data passed through health data repository firewalls.**

Index Terms—**Clustering, ICU, Mortality Prediction, Regression Analysis, Summarized Data**

I. Introduction

Clinical decision support tools based on patients who are similar to the current patient [2] might help physicians deal with the cognitively complex, intense, and interrupt driven tasks (e.g., [3]) that they are often faced with. Chignell et al. [1] proposed the use of cluster analysis to group patients into non-confidential patient types. They demonstrated the feasibility of this approach by clustering the MIMIC II database of intensive care unit (ICU) records into a set of patient types, and it seems likely that similar analyses could be carried out on many other health data repositories to develop meaningful patient types. However, there are two important criteria that summarized patient types must satisfy before they can be used in practice. First, they must be designed so that they preserve privacy and second, they must be useful, in terms of making valid predictions and providing support to clinical decision-making. Summarizations of data and applications that use them should be designed so as to minimize risks of identity and attribute disclosure (e.g., [4]). While we believe that carefully designed summarizations of data are an excellent way to preserve privacy, further discussion of the privacy implications of our approach is outside the scope of this paper. Here, we focus on the issue of the quality of predictions (concerning alive/dead outcomes of ICU patients) made by matching the current patient against a set of summarized patient types.

Death in the ICU is just one of many healthcare outcomes or variables that could be predicted based on summarized patient types, but it is an important problem that is currently being worked on, and it provides an opportunity to validate the methods that we are developing. In a clinical setting, being able to determine which patients are at increased risk of death should facilitate clinical decision-making and lead to better targeted interventions, and more effective use of rapid response teams.

The development of abstract patient types (clusters of similar patients) from confidential medical records [1] is proposed as the first step in developing applications that present summaries of similar patient types (or predictions based on those summaries) as an aid to clinical decision making. In contrast to approaches that focus only on particular prediction tasks (such as predicting death, or severity of status, in the ICU), our ultimate objective is to develop a health data search engine that can support useful predictive analytics for a wide range of queries/problems (e.g., suggesting diagnoses, predicting length of time to the next emergency department visit, and so on). While our motivation for this work originally came from emergency medicine [3], and more recently intensive care, summarized data may be useful in supporting decision making in many areas of healthcare. In carrying out the research reported below we hypothesized that summarized data derived from each cluster (patient type) could be at least as informative as case level data (individuals). Our goal in the research reported here was to evaluate how well data that is 1) blindly clustered without any regard for the prediction task to be used, and then 2) summarized, can support subsequent prediction (vs. predictions made using standard techniques

Submitted for review on April 19, 2014. This work was supported, in part, by a Google Faculty Research Award, and by a Discovery Grant from the National Science and Engineering Research Council of Canada.

Mahsa Rouzbahman (email: mrouz@mie.utoronto.ca) and Mark Chignell (email: chignell@mie.utoronto.ca) are both with the Department of Mechanical and Industrial Engineering at the University of Toronto, Toronto, Ontario M5S 3GA, Canada (e-mail: author@ boulder.nist.gov).

[1] ICU: Intensive Care Unit

applied to case level data).

This paper begins by reviewing relevant literature on clinical decision support, on predicting values in healthcare data, and on predicting death in ICUs in particular. We then provide an overview of our method for summarizing confidential data (see [1] for more details on the method). Using a set of published ICU data we then retrospectively predicted if patients in each cluster died or not using reported death status (alive or dead) as the dependent (criterion) variable. Predictions based on linear regressions carried out with summarized data were compared with regression analyses (linear and logistic) and discriminant analysis carried out on the case-level (non-summarized data). Our overall finding is that prediction based on the summarized clusters (patient types) is better than prediction based on the non-summarized original data (for this ICU data set). We conclude by noting the likely utility of summarized data as a basis for health data search engines, and for associated clinical decision support tools that can predict unknown features associated with a case of current interest.

II. Background

Vinod Khosla (co-founder of Sun Microsystems) claimed in late 2012 that "Technology will replace 80% of what doctor's do" [5]. He argued that much of what doctor's do could be better done with computers: "doctors aren't supposed to just measure. They're supposed to consume all that data, consider it in context of the latest medical findings and the patient's history, and figure out if something's wrong. Computers can take on much of that diagnosis and treatment and even do these functions better than the average doctor (while considering more options and making fewer errors)." While it might be some time before this envisioned future comes to pass, where software agents act as physicians, there seems little doubt that systems that support clinical decision-making appropriately can enhance treatment outcomes today. Early clinical decision support systems (CDSS) were often linked to computerized provider order entry (CPOE) systems and research evidence indicated that the addition of clinical decision support (CDS) enhances healthcare quality and efficiency [6].

Later research addressed the issue of what features of CDSS led to improved performance. Garg et al. [7] found that CDSS that automatically prompted users performed better than those that required users to activate the system. Kawamoto et al. [8] identified four features that independently predicted improved clinical practice: automatic provision of decision support as part of clinician workflow; provision of recommendations rather than just assessments; provision of decision support at the time and location of decision making; computer based decision support.

CDS can be based on best available clinical evidence or guidelines, derived from evidence-based medicine [9]. However, when Shojania et al. [10] carried out a systematic review of point of care computer reminders on physician behaviour they found that such reminders produced smaller improvements (in the order 4-5%) than those generally expected from the implementation of computerized order entry and electronic medical record systems.

CDS can also utilize predictions based on data mining of previous cases as has been proposed for pathology ordering [11]. While interest in data mining of health care records to support decision making is high, the topic is still in its infancy, with little agreement as to which of the many available data mining methods should be used, and how the results should be integrated into practice.

A useful starting point for the design of CDS based on data mining of health records is to look at how physicians currently make decisions. This activity may provide clues as to how to integrate decision support more effectively into their practice. Medical decision makers are influenced by previous cases that they have seen, and this is particularly true for older doctors, who will generally be more comfortable with reviewing similar cases than they will be in using search technologies [12]. Presentations of data about similar patients can be a form of clinical decision support [2]. However in environments like emergency medicine, where there may be numerous interruptions, and high workload [3], emergency (or intensive care) physicians would likely not have time to search through records of similar patients, even if they had real-time access to those records. While information about patients that are similar to the current one can be done through a background search process [13] time is still required to go through the similar patient summaries and consider them in the context of the current case. An alternative approach is to make predictions based on what is known about the patients that are judged to be similar to the current case. These predictions could then be presented to the physician as recommendations, or suggestions about decisions or actions to be taken, or probabilities of outcomes associated with particular actions.

Gottlieb et al. [14] provided an example of decision-making based on patient similarities. They used demographic, initial blood and electrocardiography measurements, as well as medical history of hospitalized patients from two independent hospitals to predict eventual discharge diagnoses. Their method involved combining data from two hospitals and it also utilized a data set of 55 million associations between 5.8 million patients and 1,125 third level discharge ICD codes.

While impressive results have been obtained in using health data repositories to make diagnostic and treatment decisions, they have generally involved considerable time, resources, and privileged access to confidential health data. Is it possible to use summarized, non-confidential data to make novel and useful predictions about unknown features, based on easily assessed statistical properties of the data, within a matter of seconds and without using detailed medical knowledge?

In order to demonstrate the value of summarized patient types in making predictions, we chose to look at the problem of deterioration and death in the ICU. ICU patients typically need constant and close monitoring, leading to large amounts of data on each patient. Prediction of who is likely to die in the ICU, or after leaving the ICU, is important both in preventing premature death and in allocating resources (such as rapid response teams) more efficiently [15]. Prediction of death in

the ICU has been a focus of research for some time. For instance, the APACHE[2] scoring system was developed for assessing risk of death in intensive care [16]. Ohno-Machado et al. [17] summarized much of the subsequent work on data mining and modeling of intensive care outcomes. Examples of this work included [18], which developed a logistic regression model to predict risk of death for children (under age 16) in an intensive care unit, and [19] which applied univariate and multivariate logistic regression analysis to find independent predictors of death in patients with acute type A aortic dissection. More recently, Hug [20] developed a real-time acuity score suitable for continuous risk assessment of ICU patients. Joshi and Szolovits [21] used an unsupervised learning approach to visualize the status of multiple organ systems in real time. While their technique was unsupervised, it was focused on the status of organ systems and thus specialized to some extent.

In a recent study, multivariable logistic regression was used to predict "out of intensive care unit" cardiopulmonary arrest, or death. The authors considered vital signs, laboratory data, physician orders, medications, floor assignment, and the Modified Early Warning Score (MEWS) as predictors in this study [15]. Other studies have made predictions based on similar patient data. Gotz et al. [22] presented an approach to extract a cohort of patients that are similar to a target patient. Ebadollahi et al. [13] predicted the trajectory of a patient's physiological data based on temporal trends in similar patients.

In summary, there is considerable interest in data mining of health records to support clinical decision making, but there are also obstacles. In addition to concerns about privacy, there is a need for techniques that can support decision-making in real-time, and that do not require large, and costly efforts to develop associated knowledge bases or ontologies (e.g., [23]). Thus our focus on this paper will be on a relatively assumption-free form of lightweight data mining and generalized prediction that could ultimately be conducted in real time as a Web service.

III. Clusters Of Similar Patients

In conducting this research, we worked with the MIMIC II database (available from physionet.org), which is a rich database of ICU data containing a wide range of medical data, including demographics, tests, charts and medical reports.

K-means clustering was carried out on the adult patients in the data set (around 25,000 patients), varying the number of extracted clusters between 3 and 50. Prior to doing the clustering we carried out factor analysis to identify groups of highly correlated variables that were likely reflecting the same underlying factor. We addressed the issue of multi-collinearity (high inter-correlations) by replacing highly correlated variables within a factor with a unit-weighted scale that combined those variables into a single score. Since we used the Euclidean distance measure with K-means, we also standardized (using a z-transformation) the variables prior to clustering so that the analysis was not unduly influenced by differences in scale (units) between the variables. Another standard method that we used to clean up the data prior to the analysis was replacement of extreme outliers (more than five standard deviations from the mean) with less extreme values (two standard deviations from the mean). We chose a 40 cluster solution as having the best combination of a) large number of clusters, and b) also containing a sizeable group of people within each cluster. Within this solution we focused on 27 clusters that each had at least 100 patients in them. Eight of the clusters contained more than 1,000 patients, with the largest cluster having over 3,000 patients.

The cluster analysis is discussed elsewhere [1]. In this paper we focus on summaries of the clusters to carry out prediction. Figure 1 summarizes the clustering structure that was obtained by [1] (27 clusters with seven of the clusters forming a super-cluster obtained using factor analysis of the clusters). While not shown in this figure, larger clusters could be further partitioned until desired levels of size, and homogeneity within clusters, are reached. The clustering shown in Figure 1 was based on medications, lab results, chart information and some demographics and ICU details (such as length of stay).

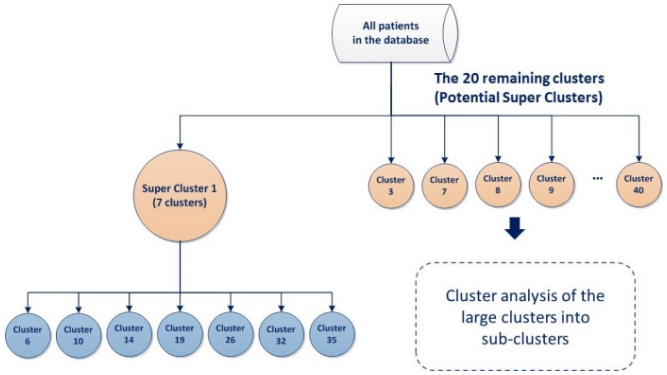

Fig. 1. Clusters of similar patients: Structure of the 27 Clusters

Chignell et al. [1] used an analysis of the difference vectors between cluster centroids to interpret each cluster based on its unique (discriminating) features. However, when using their clusters for prediction, we found it best to utilize all the features in building regression models. Clusters were summarized with the mean, median, maximum, minimum and standard deviation for the features. The feature inter-correlation matrix for patients within the cluster was also stored so that regression analysis could then be carried out on the summarized data.

How can one make predictions on the basis of summarized data? First, one needs to match the current patient against a set of previously identified clusters (patient types). This matching is based on similarity between the patient and the vector of features that constitute the summarized cluster representation.

One technique that is typically used for matching is k-th nearest neighbour assignment (e.g., [24]), but it could not be applied in our case because individual data records (neighbours) are not available with summarized data. However, patients can be matched to a cluster based on which cluster centroid they are closest to. Analogously to k-th

[2] APACHE: Acute Physiology and Chronic Health Evaluation

nearest neighbour assignment, one could do k-th nearest cluster assignment, where predictions for each of the k-th nearest matching clusters for each patient can be averaged to make predictions, or presented as alternatives if being browsed. In the case study reported below we used the computationally simpler method of making predictions based on the closest matching centroid (i.e., 1-th nearest cluster assignment). Since we only had 27 clusters this seemed reasonable and it yielded satisfactory results. Since the matrix of feature inter-correlations was stored as part of the summarization of the cluster, regression analysis could then be carried out on the cluster summaries. This works because the correlation matrix (with means and standard deviations) can be used as the starting point for linear regression analysis, yielding the same predictions (with standard multiple linear regression models) as would have been obtained with a case-level regression analysis of the non-summarized data.

IV. Prediction with Summarized Data versus Original Case Data

Patients within the Mimic II patient type clusters that we identified had significant differences in death rate, even though occurrence (or otherwise) of death wasn't one of the features used in the cluster analysis that identified the clusters. Could prediction of death be used to validate the use of summarized data, as a substitute for the original case level data, in making a clinically relevant prediction?

TABLE I
Cluster details (number of patients and dead percentage)

Cluster ID	Number of Patients	Dead Percentage
3	182	47
7	1972	57
8	262	54
13	132	16
15	4159	21
17	311	32
18	363	45
22	3036	31
38	758	88
39	123	27
23	199	32
9	2803	39
31	186	27
34	203	31
25	483	14
36	223	46
27	130	31
37	103	NA
29	193	56
40	103	53

To compare prediction between summarized and original case data we applied multiple linear regression analysis with outcome of each patient (0=live, 1=dead) as the criterion. Typically, logistic regression analysis or discriminant analysis are used with a dichotomous criterion, and we used those methods with the case level (non-summarized) data. However, for data summarized with correlations, linear regression analysis is feasible, but not the other two methods. For regression analysis the predicted occurrence of death was treated as a continuous variable (probability of death) that varied between 0 and 1. To obtain actual predictions of status, the predictions on the continuous interval (0,1) were then rounded to the nearest integer value (0 or 1).

Table I shows number of patients and percentage of patients who died within each cluster. We used the four largest clusters (7, 9, 15, and 22, highlighted in the table) in our initial analyses below.

A. Prediction of Death Methodology

To predict occurrence of death with original case data, we considered the (actual) occurrence of death as the criterion, and the other features as predictors. We analyzed the data with both logistic regression and discriminant analysis. For each of those procedures we first used the procedure on the entire data set, selected those participants who were in the cluster of interest, and then calculated accuracy based on the amount of agreement between the predictions made (over all the participants in the cluster) and the actual death status for those people. Thus only one discriminant analysis and one logistic regression analysis were run, with different sets of participants being used to calculate the accuracy score depending upon which cluster was under consideration. In addition, for the sake of later comparability with summarized prediction, we also ran a linear regression analysis on the original case data where the regression equation was calculated based on all the data and then that equation was fitted to whichever pool of participants was being considered (i.e., the participants of a particular cluster). For instance, to calculate the case level regression accuracy for cluster 7 we predicted the death status for all the participants in cluster 7 based on the regression equation calculated over the entire data set (across all the clusters), and compared the resulting predictions with actual death status.

Accuracy was calculated as the number of correct predictions, i.e., alive when predicted alive, dead when predicted dead, divided by the total number of predictions within the cluster.

V. Results

We carried out the analyses, described above, for each of the four largest clusters (clusters 7, 9, 15, and 22). Tables II to V summarize the results that were obtained from the resulting four analyses.

TABLE II
Cluster 7: Accuracy of the methods

Cluster 7	Accuracy
Regression analysis on summarized data	68.66
Linear regression analysis case level	67.39
Logistic regression case level	70.99
Discriminant analysis case level	71.29

In cluster 7, the logistic regression and discriminant analysis were both more accurate than the regression models, with the summarized regression being a little more accurate than the case level regression.

For cluster 9, prediction on the summarized data was over six percent more accurate than regression analysis on the case level data, and the summarized regression predictions was also close to three percent more accurate than the corresponding predictions for the discriminant analysis and the logistic regression.

TABLE III
Cluster 9: Accuracy of the methods

Cluster 9	Accuracy
Regression analysis on summarized data	75.31
Linear regression analysis case level	68.99
Logistic regression case level	72.45
Discriminant analysis case level	72.49

For cluster 15, prediction with summarized regression was between one and two percent more accurate than prediction with the other methods.

TABLE IV
Cluster 15: Accuracy of the methods

Cluster 15	Accuracy
Regression analysis on summarized data	81.101
Linear regression analysis case level	79.03
Logistic regression case level	79.85
Discriminant analysis case level	79.87

For cluster 22, the summarized regression prediction accuracy appeared slightly (around 1.5%) better than corresponding accuracy for case level regression. However, summarized prediction accuracy was marginally worse than the accuracy for logistic regression and discriminant analysis.

TABLE V
Cluster 22: Accuracy of the methods

Cluster 22	Accuracy
Regression analysis on summarized data	74.20
Linear regression analysis case level	72.72
Logistic regression case level	74.93
Discriminant analysis case level	75.09

In summary, accuracy of the summarized regression predictions was sometimes better, sometimes comparable, and sometimes worse than logistic regression and discriminant analysis in the four clusters considered here.

Since summarized prediction outperformed the single regression model fitted to the entire (case level) data set, it would seem to be advantageous to use different regression models within different clusters and this appears to be the case (since the list of significant features/predictors varied considerably across the four different regression analyses used for the four clusters highlighted here). For instance, while one lab variable was a significant predictor in three of the four clusters, a second lab variable was only a significant predictor in cluster 9 and a third lab variable was only a significant predictor in cluster 7.

Inspection of Table VII suggests that relationships between the EHR[3] features and death status varied between the clusters. Could summarized prediction take advantage of this variation and out-predict logistic regression and discriminant analysis on case level data, when more clusters where included in the analysis?

When the comparison was repeated but for the people in all 27 of the clusters (Table XIII), prediction made using only the summarization data was almost 5% more accurate than that of logistic regression, or discriminant analysis, using the original case level data.

TABLE VI
All 27 Clusters: accuracy of different methods

All Clusters	Accuracy
Regression analysis on summarized data	81.87
Linear regression analysis case level	75.40
Logistic regression case level	77.2
Discriminant analysis case level	77.2

VI. Discussion

When interpreting the results presented above it is important to keep in mind how different our methodology was from the best practices typically used in healthcare data mining. The cluster analysis was carried out on a publicly available data set and we had no information about the data set other than the documentation provided. The authors of this paper had no medical training to guide them in doing the analysis and the only input from medical experts came when two physicians reviewed the obtained clusters to provide a sanity check on whether they seemed to make some sense. The cluster analysis was performed using a standard technique available in a widely distributed statistical package and there was no supervision of the clustering processing other than choosing how many clusters were to be extracted. Only one clustering method was used and the cluster analysis was done in a single shot (but with a varying number of partitions). It is almost certain that the cluster analysis we did could be improved upon if more effort and expertise were applied to it. However, the methods that we used were designed to be easily automated, so that they could be carried out automatically and efficiently within a health data repository that communicated with a search engine/web service.

Summarized prediction of death within a cluster was not noticeably more accurate than using logistic regression, and discriminant analysis, with the case level data. However, when using prediction across multiple clusters, summarized prediction outperformed logistic regression, and discriminant analysis, carried out on the raw data. When we looked at only the four clusters pooled together, the advantage in accuracy of prediction for summarized regression (versus logistic regression and discriminant analysis) was a little less than 2%. However, this advantage rose to almost 5% when all 27 clusters were considered together.

[3] EHR: Electronic Health Record

How can predictions based on summarized data outperform predictions based on the original, un-summarized data? Assigning people to clusters likely boosts accuracy because differentiation of patients into clusters effectively partitions the high dimensional multivariate space of predictor variables into subspaces that are likely to be governed by different regression models within each subspace. If regression relationships differ within different regions of the multivariate space, then separately fitting different regression models within different subspaces should lead to more accurate fitting than trying to fit a single model over the entire high dimensional space.

The two-step process of clustering followed by regression used in this paper was originally described by [25], who referred to that process as piecewise regression. We prefer the term cluster-boosted regression (CBR) because the "piecewise" descriptor seems more applicable to low-dimensional spaces (and two-dimensional line fitting in particular) than to the high-dimensional spaces that are typical of healthcare data mining applications. The major difference in our approach was that we used k-means analysis for clustering rather than the hierarchical clustering used by McGee and Carleton. We prefer the k-means clustering method because it is a simple and very efficient method that is widely available in many packages include the open source R statistical package. Most importantly though, it partitions the high dimensional feature space of patient electronic health records into roughly spherical regions or subspaces within which specialized regression models can be fitted. Within a particular subspace some variables may be "flattened", i.e., have little variability (e.g., the patients within a cluster may have a similar length of stay). Additionally, relationships (correlations) between variables may also differ from one subspace (cluster) to another, leading to different regression models applying within different subspaces. Since the goal of the clustering is to enhance predictions made with regression models, it doesn't matter if the clustering solution is a completely accurate representation of some underlying conceptual structure behind the data. By partitioning the data into reasonably homogeneous subspaces, regression models are specialized to those subspaces and will tend to fit better.

Given that the clusters used in the present study were formed in an ad hoc fashion by researchers with little medical expertise we believe that the result that we obtained will likely be replicated in summarized prediction in other healthcare contexts. While this expectation needs to be validated in future research, it seems likely that CBR should make summarized regression/prediction an effective strategy in a wide variety of contexts. Although summarized prediction is not always competitive within a single cluster it appears to become more accurate as more people (and clusters) are included in the prediction. The 82 percent accuracy in predicting death across the 27 clusters in our sample is no doubt poor when compared to knowledge-based techniques, and to techniques that take into account temporal effects and known causal relationships. Presumably the accuracy of the generalized prediction method used here could be boosted with better knowledge, or by supplementing predictions with knowledge-based heuristics (e.g., in one paper we reviewed risk was adjusted based on which floor of the hospital a patient was on).

As our results show, even an unsupervised clustering process, run by medically unskilled data analysts, can identify meaningful subspaces. Within such subspaces summarized regressions can outperform analyses carried out on the original (unpartitioned) feature space using case level data.

VII. Conclusion

Across multiple clusters, summarized regression outperformed case level analyses rather handily in this study. We attribute this to the fact that the derived clusters of patient types partitioned the multivariate feature space effectively and captured important statistical relationships and differences that boosted the regression results.

While not a focus of this paper, we believe that with sufficiently large clusters of patient types being summarized (and with the addition of a certain amount of randomization, if found to be necessary), privacy issues arising from using summarized version of patients' data should be relatively minor. Without detailed knowledge of the data and the clustering process there is no way back from correlation, mean and standard deviation to individual data records. In contrast to tables of data returned in response to queries, clustered summaries returned in response to queries are much harder to deconstruct since clusters are dependent on random initial partitions, and exactly which records are included in the clustering. Differing clustering methods can also be intermixed, making it even harder for malicious outsiders to try and infer individual data since they will have no way to model the details of the clustering process. This opacity of summarized prediction is likely to be even more evident if only predictions for individual cases are provided and the predictive analytics is actually carried out inside the firewall.

The approach to healthcare data mining introduced here offers re-use of health data in clinical decision support while providing methods that preserve privacy and offer better prediction. We hope that other researchers will be able to confirm and extend the present results with respect to the benefits of summarized health data in building predictive analytics and healthcare applications based on non-confidential health data. We see the present research as an encouraging step towards a general search engine that can accept queries and then pass them on to a large number of health data repositories that would then return matching clustered summaries suitable for prediction and clinical decision support. While the appropriate user interface and interaction paradigm for associated clinical decision support tools have yet to be elaborated, a first step may involve predicting unknown features associated with a case of current interest and flagging them as possibilities in the patient's data record.

ACKNOWLEDGMENT

We would like to thank Roger Mark and his colleagues for making the MIMIC II database available to researchers like us. This research would not have been possible without it. We also thank Peter Szolovits for discussing the problem of predicting death in the ICU with us. Portions of this research were funded by a Google Faculty Research Award to the second author and by an NSERC Discovery Grant to the second author.

REFERENCES

[1] M. Chignell, M. Rouzbahman, R. Kealey, E. Yu, R. Samavi and T. Sieminowski, "Development of Non-Confidential Patient Types for Use in Emergency Medicine Clinical Decision Support" Security & Privacy, IEEE Volume:PP , Issue: 99, 2013.

[2] L. W. C. Chan, "Machine Learning of Patient Similarity,"IEEE International Conference on Bioinformatics and Biomedicine Workshops, 2010.

[3] E.Yu, R. Kealey, M. Chignell, J. Ng and J Lo, "Smarter Healthcare: An Emergency Physician View of the Problem, "In M. Chignell, J. Cordy, J. Ng., and Y. Yesha, (Eds). The Smart Internet: Current Research and Future Applications. Lecture Notes in Computer Science, 6400. Berlin: Springer, 2010.

[4] K. El Emam, "Methods for the de-identification of electronic health records for genomic research," Genome Medicine, 3(4), 25, 2011.

[5] D. Liu, "Vinod Khosla: Technology Will Replace 80 Percent of Docs," The health care blog, 2012. Available at: http://thehealthcareblog.com/blog/2012/08/31/vinod-khosla-technology-will-replace-80-percent-of-docs/.

[6] D.L. Hunt, R.B. Haynes, S.E. Hanna, et al., "Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review," JAMA 1998;280:1339–46.

[7] A.X. Garg, N.K. Adhikari, H. McDonald, M.P. Rosas-Arellano, P.J. Devereaux, J. Beyene, J. Sam and R.B. Haynes, "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes," JAMA: the journal of the American Medical Association 293, no. 10 (2005): 1223-1238.

[8] K. Kawamoto, C.A. Houlihan, E.A. Balas and D.F. Lobach, "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success," Bmj, 330(7494), 765, 2005.

[9] Straus, S. E., Glasziou, P., Richardson, S., & Haynes, B. (2010). Evidence-Based Medicine: How to Practice and Teach it (Fourth Edition). London: Churchill Livingstone Elsevier.

[10] K.G. Shojania, A. Jennings, A. Mayhew, C.Ramsay, M. Eccles and J. Grimshaw, "Effect of point-of-care computer reminders on physician behaviour: a systematic review," Canadian Medical Association Journal, 182(5), E216-E225, 2010.

[11] Z. Y. Zhuang, L. Churilov, F. Burstein and K. Sikaris, "Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners," European Journal of Operational Research, 195(3), 662-675, 2009.

[12] N.K. Choudhry et al., Systematic Review: "The Relationship Between Clinical Experience and Quality of Health Care," Annals of Internal Medicine, No. 142, Issue 4, 2005.

[13] S. Ebadollahi, J. Sun, D. Gotz, J. Hu, D. Sow and C. Neti, "Predicting patient's trajectory of physiological data using temporal trends in similar patients: a system for near-term prognosis, " Proceedings of AMIA 2010.

[14] A.Gottlieb, G. Y. Stein, E. Ruppin, R. B. Altman and R. Sharan, "A method for inferring medical diagnoses from patient similarities, " BMC medicine, 11(1), 194, 2013.

[15] C.A. Alvarez, C. A. Clark, S. Zhang, E. A. Halm, J.J. Shannon, C. E. Girod, L. Cooper and R. Amarasingham, "Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data," BMC medical informatics and decision making, 13(1), 28, 2013.

[16] J.R. Le Gall, P. Loirat, A. Alperovitch, P. GLASER, C. GRANTHIL, D. Mathieu and D. Villers, "A simplified acute physiology score for ICU patients," Critical care medicine, 12(11), 975-977, 1984.

[17] L. Ohno-Machado, F.S. Resnic, M.E. Matheny "Prognosis in critical care," Annual Review of Biomedical Engineering, 8, 567–599, 2006.

[18] F. Shann, G. Pearson, A. Slater and K. Wilkinson, "Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care," Intensive care medicine, 23(2), 201-207, 1997.

[19] R.H. Mehta, T. Suzuki, P.G. Hagan, E. Bossone, D. Gilon, A. Llovet and K.A. Eagle, "Predicting death in patients with acute type A aortic dissection," Circulation, 105(2), 200-206, 2002.

[20] C. Hug, "Detecting Hazardous Intensive Care Patient Episodes Using Real-time Mortality Models," PhD thesis, MIT, 2009.

[21] R. Joshi and P. Szolovits, "Prognostic Physiology: Modeling Patient Severity in Intensive Care Units Using Radial Domain Folding," Proceedings of AMIA, 2012.

[22] D. Gotz, J. Sun, N. Cao and S. Ebadollahi, "Visual Cluster Analysis in Support of Clinical Decision Intelligence," In AMIA Annual Symposium Proceedings (Vol. 2011, p. 481). American Medical Informatics Association (2011).

[23] M. Samwald, R. Freimuth, J. S. Luciano, S. Lin, R. L. Powers, M. S. Marshall and R. D. Boyce, "An RDF/OWL Knowledge Base for Query Answering and Decision Support in Clinical Pharmacogenetics," Studies in health technology and informatics, 192, 539, 2013.

[24] R.O. Duda, P. E, Hart and D. G. Stork, "Pattern classification," John Wiley & Sons, 2012.

[25] V.E. McGee and W.T. Carleton, "Piecewise regression," Journal of the American Statistical Association, 65(331), 1109-1124, 1970.

Mahsa Rouzbahman is currently a PhD candidate of Mechanical and Industrial Engineering at the University of Toronto. She received a M.Sc. degree in industrial engineering from University of Tehran. Her main fields of interest are human factors research, user interface design for health care environments, clinical decision support systems and data mining of medical records. Contact her at mrouz@mie.utoronto.ca.

Mark Chignell is a Professor of Mechanical and Industrial Engineering at the University of Toronto, and Director of the Knowledge Media Design Institute at the University of Toronto. He's interested in making people smarter and more effective through better design of user interfaces and applications. He has a Ph.D in Psychology from the University of Canterbury in New Zealand and a Masters in Industrial and Systems Engineering from Ohio State. Contact him at chignell@mie.utoronto.ca.