Array 8 (2020) 100049

Contents lists available at ScienceDirect

Array

journal homepage: www.elsevier.com/journals/array/2590-0056/open-access-journal

Using machine learning to identify top predictors for nurses' willingness to report medication errors

Renjie Hu [a,*], Amany Farag [b], Kaj-Mikael Björk [c,d], Amaury Lendasse [c,e]

[a] Department of Industrial and Systems Engineering, The University of Iowa, Iowa City, USA
[b] College of Nursing, The University of Iowa, Iowa City, USA
[c] Arcada University of Applied Sciences, Helsinki, Finland
[d] Hanken School of Economics, Helsinki, Finland
[e] Department of Information and Logistics Technology, College of Technology, University of Houston, Houston, USA

ARTICLE INFO

Keywords:
Variable selection
Data visualization
Dimensionality reduction
Extreme learning machines
Self-organizing maps
Medication error reporting

ABSTRACT

This paper presents a novel methodology to analyze nurses' willingness to report medication errors. Parallel Extreme Learning Machines were applied to identify the top interpersonal and organizational predictors and Self-Organizing Maps to create comprehensive visualization. The results of the data analysis were targeted to improve the likelihood of nurses reporting of medication errors. ELMs are accurate by extremely fast prediction models. Self-Organizing Maps enable us to perform non-linear dimensionality reduction to get an accurate visualization of the selected variables. Combining both techniques reduces the curse of dimensionality and improves the interpretability of the visualization.

1. Introduction

According to the seminal institute of medicine report, medical errors ranked as the leading cause of death in the United States [1], and recently climbed to the third leading cause of death [2]. It is estimated that about 44,000 to 98,000 people die annually from medical errors [1]. These numbers are higher than deaths from breast cancer, AIDS, and car accidents combined. Medication errors are the most frequently occurring medical error in healthcare settings [3]. Unfortunately, serious life threatening errors are usually reported, but the majority of other medication errors are not [4].

Medication delivery is a complex multi-stage process that involves several healthcare professionals [5]. Medication errors could occur at each step of the medication process [6], with 38% of errors occurring at the administration phase [7]. Nurses spend about 40% of their time administering medications, and by virtue of their position represent the last safety defense to intercept errors before reaching their patients [8]. Most hospitals relay on nurses to report medication errors, in some cases nurses might be the witness or committer of medication error [9].

Medication error reporting is a voluntary process [10]. Reviewing and analyzing medication error report, known as incident report, provides healthcare administrators and safety officers with opportunities for understanding error root causes and subsequently design interventions to prevent subsequent errors [11–13]. However, having less than 5% of errors reported, makes developing a proper medication error intervention a tough challenge [14]. Fear of blame, punishment, humiliation, retaliation from managers and/or peers were some of the reasons deterring nurses from reporting errors bib15[9,15]. Mayo and Duncan (2004) [9] argued that all efforts of healthcare administrators, policy makers and scholars to create effective medication errors reporting systems, could fail if nurses remain unwilling to report errors. Limited information exist about organizational and interpersonal variables needed to motivate nurses to report medication errors. Therefore, the purpose of the three nursing studies is to identify interpersonal and organizational variables influencing nurses' willingness to report medication errors.

Given the complexity of healthcare organizations, and the non-linear relationships among aspects of healthcare systems, interpersonal relationships, and nurses willingness to report medication errors, using traditional regression approaches might often lead to inaccurate results [16,17], and limited interpretability [18,19] which will challenge proposing appropriate improvement interventions. Another challenge is that: generally there are large number of the predictor variables, so the possible combinations of the variables (predictors) are extremely large, which makes it difficult to identify which combination is the most effective for predicting nurses' willingness to report.

In this paper, a new method has been proposed to analyze predictors

* Corresponding author.
E-mail address: rh2703@columbia.edu (R. Hu).

https://doi.org/10.1016/j.array.2020.100049
Received 20 December 2019; Received in revised form 24 August 2020; Accepted 11 October 2020
Available online 9 November 2020

of nurses' willingness to report medication errors and visualize the non-linear relationship between the variables and the nurses' willingness. To identify the top predictors a new parallel variable selection method is presented with Extreme Learning Machines (ELMs) [20–23]. ELMs is accurate by extremely fast prediction models [20], therefore, it is possible with them to test a very large number of possible variables.

The selected variables are visualized by comprehensive Self-Organizing Maps (SOMs) [24–27]. Self-Organizing Maps are performing non-linear dimensionality reduction [24] to get an accurate visualization of the data, which significantly enhances the interpretability of the selected predictors, and reveals the determinant factors for nurses' willingness to report medication errors, therefore guides the hospital to conduct intervention to improve the likelihood of receiving medication error reports from nurses.

The paper is organized as follows: in Section 2 the detailed problem description is provided. Section 3 discusses about the process of identifying the top predictors. Visualization technique is introduced in Section 4. The complete experiment settings are presented in Section 5. Final results are included in Section 6. Conclusions and the future works are in Section 7.

2. Nurses error report dataset

Three survey data were collected from three funded nursing projects. The projects were conducted to identify interpersonal and organizational predictors of nurses' willingness to report medication errors.

Study Design: non-experimental cross-sectional design was used in the three projects. Setting: data for the three projects were collected from Registered Nurses (RN) working in general medical surgical units, critical care units, nursing homes, and Emergency Departments (ED) from multiple hospitals in one Midwestern state.

Study Sample: Full-time, part time, and as needed base nurses (PRN), with no leadership responsibilities regardless their age, gender, years of experience were invited to the studies. Nurse managers, nurse practitioners, and those who are not involved in passing medication were excluded from the studies.

Study Measures/Instruments: Nurses' demographics were measured using a PI developed demographic questionnaire that includes: nurses' age, gender, total years of experience, highest nursing degree, type of unit, years of experience with the current nurse manager, years of experience in the current unit, number of hours worked per week and type of shift (morning, evening, or rotating). In addition to the previous measurements the surveys included further measurements including, Interpersonal variables, Organizational variables (leadership and unit climate) and Outcome variables of nurses' willingness to report medication errors.

2.1. Data preparation

Each survey data is collected in a separate.csv file. The variables in the dataset are corresponding to the questions from the survey, and the values of the variables are the subjects' answers to the questions, which are coded in the numerical Likert scales. The name of the variables are coded in two parts: the abbreviation of the survey section name followed by the question number. For example: variable *LSHPQ1* means question 1 in the section of Nurse manager's leadership style, representing the question: My unite manager provides me with assistance in exchange for my efforts. Table 1 shows the correspondences among the Measurements, and the Number of Questions in this category. Three distinct outcomes variables are asked in the survey to measure the nurses' willingness to report medication errors in different scenarios.

The samples with missing values in the dataset have been omitted. The final experimental data contains 68 variables and 328 samples/nurses.

There are three different outcome scenarios corresponding to three distinct outcome variables, which are measured in the survey as follows:

Table 1
Survey measurements summary.

Measurements	Number of Questions
Nurse manager's leadership style	28
Warmth and belonging climate	11
Organizational trust	12
Nurses Basic information*	12
Hospital error reporting system	5
Willingness to report in scenario 1	1
Willingness to report in scenario 2	1
Willingness to report in scenario 3	1

Note*: All category variables are converted to numerical values by label encoder.

ERREPQ1: When a mistake is made, but caught and corrected before affecting the patient, how likely are you to report this error?

ERREPQ2: When a mistake is made, but has no potential harm to the patient, how likely are you to report this error?

ERREPQ3: When a mistake is made that could harm the patient, but does not, how likely are you to report this error?

Table 2 presents the math notations used in this paper.

To summarize, the training data is $\mathbf{X} \in \mathbb{R}^{328\times 68}$. Three outcome variables are $\mathbf{Y}_1$, $\mathbf{Y}_2$, $\mathbf{Y}_3 \in \mathbb{R}^{328\times 1}$, corresponding to the three distinct outcome scenarios: ERREPQ1, ERREPQ2, and ERREPQ3.

3. Identify top predictor variables with parallel ELMs

The survey data after pre-processing had 68 variables in total. Such multi-variable data is usually called high-dimensional data, which can be challenging to interpret and to analyze due to the curse of the dimensionality, data redundancy and noises [28–30]. Any model built upon high-dimensional data directly usually suffers from the poor generalization performances [28]. More importantly, such high-dimensional data is also limiting the analysis model's interpretability [29,31]. Last but not least, it is difficult to create visualization of multi-dimensional data, which is an effective way of presenting the comprehensive data [29,32], enhancing the data analysis, and guiding the future actions. In this study the visualization plays an important role of guiding the management of hospitals to improve the willingness of nurses to report medication error. For the purpose of increasing the interpretability, and obtain comprehensive visualization, it is critical to conduct proper variable selection to reduce the dimensionality of the data and identify top predictor variables.

A fast parallel variable selection model is proposed by using Extreme Learning Machines [20–22] incorporated in a wrapper variable selection mechanism [33], to identify top predictors of nurses' willingness to report the medication errors. Extreme Learning Machines are accurate,

Table 2
Math notations.

Symbols	Description
$\mathbf{Y}$	All Outcome Variables
$\mathbf{Y}_i$	Outcome Variable i
$\mathbf{X}$	Predictor Variables for all the samples
$\mathbf{x}_i$	*ith* sample from $\mathbf{X}$
$\mathbf{S}_k$	Randomly selected k variables from $\mathbf{X}$
$\mathbf{W}$	ELM Input Layer Weights
$\mathbf{b}$	ELM Input Layer biases
$\mathbf{L}$	Number of Hidden Neurons in ELM
$\mathbf{h}$	ELM Hidden Neurons output
ϕ	Non-linear transformation function in ELM
β	ELM Output Layer Weights
$\mathbf{c}_s$	Best Matching Unit in SOM
α	Learning Rate in SOM
σ_λ	Neighborhood Function in SOM
d	Distance Function in SOM

extremely fast and non-linear prediction models [20,21], therefore, it is possible with them to test a very large number of possible combinations of variables in parallel and obtain superior results than the traditional linear models bib34[16,19,34,35]. The ELM model is explained in Section 3.1. The wrapper variable selection method is explained in Section 3.3. In the variable selection process, the variables are evaluated by the R^2 value of each ELM model, this is explained in Section 3.2.

3.1. Extreme Learning Machine

Extreme Learning Machine in Ref. [23,36] as important emergent machine learning techniques, are proposed for training Single-hidden Layer Feed-forward Neural Networks (SLFNs) [20,22,37–40].

In contrast with the traditional Feedforward Neural Networks (FNNs), which generally are trained by the well-known backpropagation (BP) algorithms, in ELM, the wight for the hidden layer are randomly initiated and then fixed without iteratively tuning. Then commonly used activation functions are applied on the hidden neurons. The only parameters learned in ELM are the weights between hidden layer and the output layer. In this way, the parameters of the hidden neurons can be independent of the training data, which makes it possible for ELM to attain the near optimal generalization bound of traditional FNN. Theoretical studies as in Ref. [22,38,39] has shown that ELM has the universal approximation and classification properties.

The unique training process of ELM provides a huge leverage for the learning speed. A non-iterative solution of ELM provides a speedup of 5 orders of magnitude compared to Multilayer Perceptron ([41], MLP) or 6 orders of magnitude compared to Support Vector Machines ([42], SVM).

ELM was proposed for the fast training for the Single-hidden Layer Feed-forward Networks (Fig. 1). An SLFN has three layers of neurons: Input layer provides data features and performs no computations; Hidden layers is the only layer where the non-linear transformation happens; Output layer is linear without any transformation function and bias.

ELM is trained in two stages: First, random feature mapping. In this stage, the input data are projected randomly into a new space. ELM randomly initializes the input layer weights $\mathbf{W} \in \mathbb{R}^{d \times L}$ and biases $\mathbf{b} \in \mathbb{R}^{L}$, where L is the number of the hidden neurons. An input data $\mathbf{x} \in \mathbb{R}^{d}$ are mapped non-linearly by an activation function $\mathbf{h}_i$ onto the hidden neuron i, denoted as:

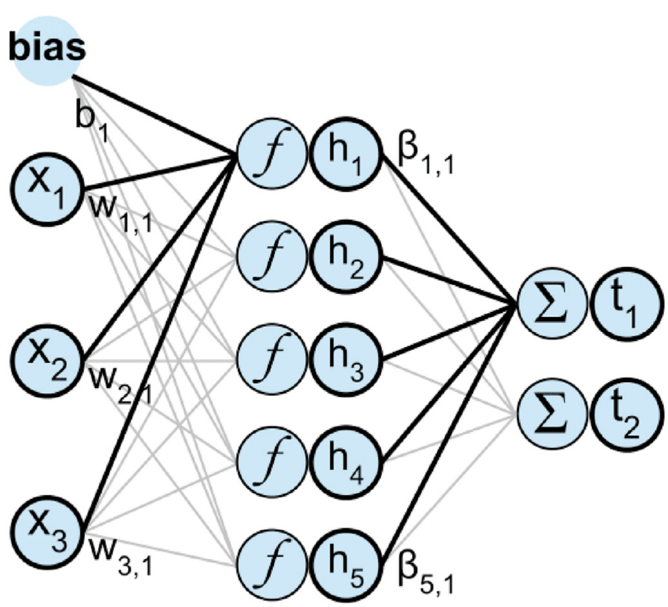

Fig. 1. Extreme learning machine.

Table 3
Typical ELM activation functions.

Hyperbolic Tangent Function	$\phi(\mathbf{w},\mathbf{x},b) = \dfrac{1-exp(-(\mathbf{x}\cdot\mathbf{w}+b))}{1+exp(-(\mathbf{x}\cdot\mathbf{w}+b))}$
Sigmoid Function	$\phi(\mathbf{w},\mathbf{x},b) = \dfrac{1}{1+exp(-(\mathbf{x}\cdot\mathbf{w}+b))}$
RBF (Gaussian)	$\phi(\mathbf{w},\mathbf{x},b) = exp\left(-\dfrac{\lVert\mathbf{x}-\mathbf{w}\rVert^2}{2b^2}\right)$

$$\mathbf{h}_i(\mathbf{x}) = \phi(\mathbf{w}_i, \mathbf{x}, b_i), \mathbf{w}_i \in \mathbb{R}^d \; b \in \mathbb{R}, \tag{1}$$

where ϕ is a non-linear piecewise continues function, generally can be selected from the following functions:

The hidden layer is not constrained to have only one type of transformation function in neurons. Different functions (see Table 3) can be used together. Some neurons may have no transformation function at all. They are linear neurons, and learn linear dependencies between data features and targets directly, without approximating them by a nonlinear function. Usually, the number of linear neurons equals the number of data features, and each of these neurons copies the corresponding feature (by using an identity $\mathbf{W} = \mathbf{I}$ and zero $\mathbf{b} = 0$).

The second stage of ELM is solving the linear system, finding the optimal $\boldsymbol{\beta}$, that minimize the following cost function:

$$MSE(\mathbf{X},\mathbf{Y}) = \min_{\boldsymbol{\beta}} \lVert \mathbf{H}\boldsymbol{\beta} - \mathbf{Y} \rVert^2, \tag{2}$$

where, $\mathbf{H}$ is the output from the hidden layer:

$$\mathbf{H} = \begin{pmatrix} \phi(\mathbf{w}_1, \mathbf{x}_1, b_1) & \dots & \phi(\mathbf{w}_L, \mathbf{x}_1, b_L) \\ \dots & \ddots & \dots \\ \phi(\mathbf{w}_1, \mathbf{x}_N, b_1) & \dots & \phi(\mathbf{w}_L, \mathbf{x}_N, b_L) \end{pmatrix}. \tag{3}$$

The cost function itself is the Mean Square Error (MSE) between the approximation of ELM ($\mathbf{H}\boldsymbol{\beta}$) and the true target value $\mathbf{Y}$. The optimal solution for the cost function, $\boldsymbol{\beta}^{\star}$ is given by:

$$\boldsymbol{\beta}^{\star} = \mathbf{H}^{\dagger}\mathbf{Y}, \tag{4}$$

where, $\mathbf{H}^{\dagger}$ is the pseudoinverse and computed as:

$$\mathbf{H}^{\dagger} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T. \tag{5}$$

Practically, the implementations of the pseudoinverse include a small regularization term $\mathbf{H}^{\dagger} = (\mathbf{H}^T\mathbf{H} + \alpha\mathbf{I})\mathbf{H}^T$.

3.2. Evaluating predictor variables with R^2 value

R^2 value can be used as an accurate measurement of variable selection performance [43]. Due to the distinct scenarios of the outcomes, the relationship between predictor variables and each outcome variables are evaluated separately, with the calculated R^2 values. This means the data are grouped into three predictor-outcome pairs for ELM training: $(\mathbf{X}, \mathbf{Y}_1)$, $(\mathbf{X}, \mathbf{Y}_2)$, and $(\mathbf{X}, \mathbf{Y}_3)$.

A R^2 value is calculated, using the MSE value from ELM the model:

$$R^2_{(\mathbf{X},\mathbf{Y}_i)} = 1 - \frac{MSE(\mathbf{X},\mathbf{Y}_i)}{Var(\mathbf{Y}_i)}. \tag{6}$$

The MSE value can also be rewrite as:

$$MSE(\mathbf{X},\mathbf{Y}_i) = \frac{1}{N}\left(\mathbf{Y}_i - \widehat{\mathbf{Y}}_i\right)\left(\mathbf{Y}_i - \widehat{\mathbf{Y}}_i\right)^T, \tag{7}$$

where, $\widehat{\mathbf{Y}}_i$ is the estimation of $\mathbf{Y}_i$ from the output of the ELM model.

The higher the R^2 value is, the more accurate prediction $\widehat{\mathbf{Y}}_i$ is. Hence, the R^2 value is used to evaluate the performance of the selected variables.

3.3. Wrapper variable selection with ELM

The wrapper variable selection method [44] targets on improving the generalization performance of the final model [45]. Although the wrapper method is more computationally expensive than the filter method [33,45], the generalization performance of the former approach is better than the later approach [33]. Multiple ELMs are built in parallel in the wrapper variable selection (see Fig. 2). Each ELM evaluates a k randomly selected variables from $\mathbf{X}$. The R^2 value is computed to measure the prediction performance of the selected k variables. The larger the R^2 values are, the better prediction of $\mathbf{Y}_i$ can be obtained from these selected variables. Therefore, variables that can produce larger R^2 values are selected over the ones produce smaller R^2 values.

The detailed variable selection process is broken down into 5 steps and the flowchart below (Fig. 3) describes the process.

Step 1. Random Variable Selection. From original dataset $\mathbf{X}$, randomly select k variables, to create a subset of the original data: $\mathbf{S}_k$. Initially, $k = 1$. The current best variable set $\mathbf{S}_k^* = \mathbf{S}_k$. The final best variable set $\mathbf{S}_k^{**} = \mathbf{S}_k$.
Step 2. ELM Training. An ELM is built upon the selected data $\mathbf{S}_k$ and the outcome variable $\mathbf{Y}_i$. The R^2 value is computed.
Step 3. Updating Current Optimal Predictor Variable Set. If the current R^2 value is higher than the previous calculation update $\mathbf{S}_k^* = \mathbf{S}_k$. Otherwise, $\mathbf{S}_k$ stays the same.
Step 4. Iterative Updation. Repeat iterative Step 1 to 3 until the stopping criteria is met. The stop criteria is defined by the maximum iteration number, which is a fairly large number. If the maximum iteration is reached, the final best variable set $\mathbf{S}_k^{**}$ is updated by $\mathbf{S}_k^*$.
Step 5. Increasing k Value. The number of variable selection is increased by one and restart from Step 1.

4. Visualize predictor variables with SOM

Self-Organization Maps [24] are used to perform visualization after the top predictors are identified. SOM is capable of revealing the nonlinear relationship between the predictors and the outcomes, which is more preferable technique than linear visualization method like PCA [25–27].

4.1. Self-Organizing Maps

SOM is a popular nonlinear dimensionality reduction tool that uses a predefined 2-D grid to capture the topology of the data in the high dimension [46,47] (see Fig. 4).

Besides the two-dimensional map representation, each point on the grid will attain a weight, or prototype, which is basically its d-dimensional representation in the original d-dimensional data space.

The grid, which consists of a rectangle including the points located on a rectangular lattice, is accompanied with randomly initialized weights for each point. Finally, after a considerable number of iterations, these weights will be updated to the points' positions in the original d-dimensional data space. In the iterative algorithm, units (or prototypes) $\mathbf{c}_s$, for $s = [1, \ldots, N]$, in which N is the number of points on a 2-D grid, are updated with the following rule:

$$\mathbf{c}_s \leftarrow \mathbf{c}_s + \alpha\sigma_\lambda(r, s)(\mathbf{x}_i - \mathbf{c}_s) \quad (8)$$

where $\mathbf{x}_i$ is the ith data point, α is a learning rate between 0 and 1, and σ_λ which is called the neighborhood function returns zeros for non-neighbors, and ones for other non-zero values for valid neighbors. In addition, d is a distance function and $r = \underset{s}{\operatorname{argmin}}\{d(\mathbf{x}_i, \mathbf{c}_s)\}$.

After the projection, according to SOM algorithm, each point $\mathbf{c}_s$, $s = [1, \ldots, N]$ on the 2-D grid is a representative of a group of points in d-dimensional data space. Basically, $\mathbf{c}_s$ is the Best Matching Unit (BMU) of a group of points in original data space.

Therefore, Self-Organizing Maps are performing a discrete nonlinear dimensionality reductions.

In order to understand the visualization, colors are used to transform the SOM into a heat map that helps understanding the importance of a given variable. Using several heat maps help analyzing the data as illustrated in the next Section.

5. Experiment setups

In this section, detailed experiment setups are discussed, including the experiment equipment, Machine Learning models' parameters, and experimental procedures.

In the first part of the experiment, the top predictor variables are identified using ELMs. Given the large amount of calculations need to perform, this part of the experiment is conducted on the University of Iowa's Argon High Performing Computing System. One CentOS-7.4 Linux Compute Node is used, with 24 processing cores and 512GB of Node Memory.

Since the three outcome questions are evaluated independently, the variable selections $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$ are running in parallel on the cluster.

Every ELM built is using 50 hidden neurons. The maximum iteration is set up as 10^8 times. The total processing time used are about 2 weeks.

One important thing to consider when conduct variable selection is

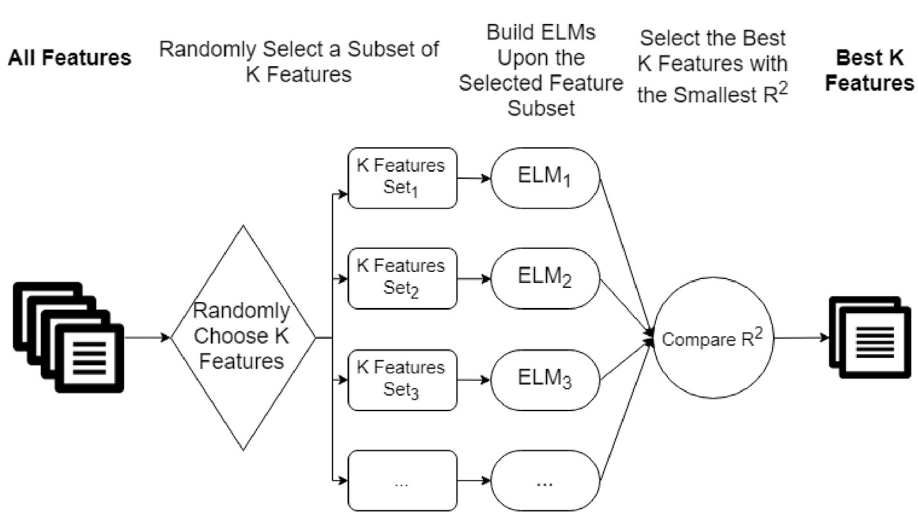

Fig. 2. Parallel ELMs for variable selection.

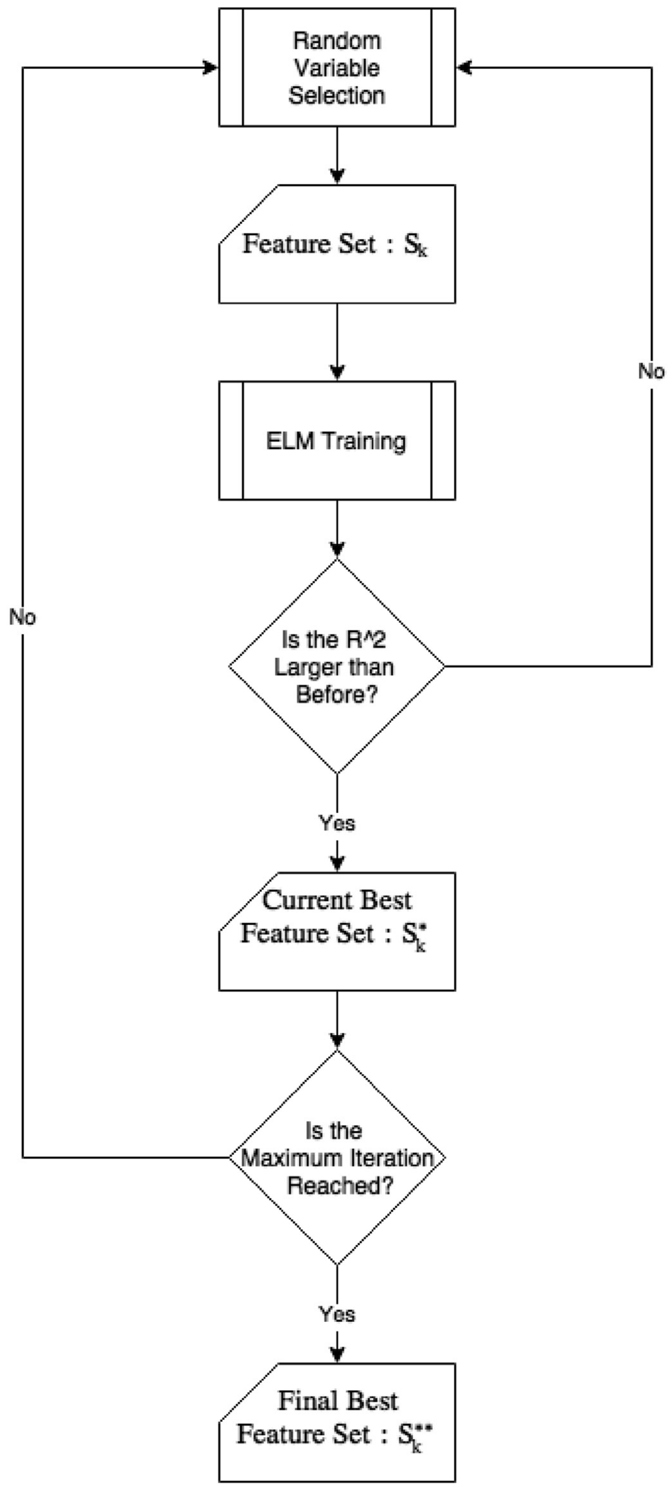

Fig. 3. Wrapper variable selection with ELM

the total possible subsets of the variables. Since there are 68 variables in total, selecting, for example, 7 variables out of 68 variables, the total possible subsets are nearly 10^8. If select 20 variables, the number of subsets could reach 8×10^{16}. Therefor, if selecting less than 7 variables, the selection is the optimal predictors; if selecting more than 7 variables, it is not possible to guarantee finding the optimal predictor variables.

Fortunately, it is not necessary to do exhaust test for all possible variable subsets to find top predictors. This is because: 1) in this study, after the optimal 3 or 4 predictors are found, the R^2 value start to gain only marginal improvements when the adding more variables. In other words, adding extra variables does not improve the prediction ability by much. Please see Section 6.1, 6.2, and 6.3. 2) to create comprehensive visualizations for human, only a few variables should be used [49,50], and some study suggests that no more than 4 variables should use [51].

Therefore, there are another set of screening rules to decide which variable sets from the selected top predictors to visualize. The two criteria used are: 1) there should not be more than 4 variables to be visualized. 2) if adding one more variable is NOT providing significant R^2 increasing, then the it shouldn't be added.

By applying the above criteria, the visualization is created by SOM, using the selected top predictors. The SOM map size is 6×8 to ensure 328 samples spreading to each units on SOM. Too large map size will lead to empty unites, while too small size will lead to overpack some unites. The learning rate α is 0.5 by default in the SOM ToolBox in Matlab, which is sufficient to guarantee quick convergence and training accuracy.

The visualization is performed on the local laptop wiht Intel(R) i7 processor and 16 GB memories.

6. Experiment results

Since the experiments for the different outcome variables are independent, the experiment results are organized by the distinct outcome variables.

6.1. ERREPQ1

ERREPQ1: When a mistake is made, but caught and corrected before affecting the patient, how likely are you to report this error?

The R^2 values for the best k variables are showing in Fig. 5.

The best variable sets $\mathbf{S}_k^{**}$, where $k = 1, 2, \ldots 20$, for $\mathbf{Y}_1$ are selected by ELMs. The following Table 4 lists selected variable names for $k = 1$ to $k = 5$.

R^2 value increases significantly at $\mathbf{S}_4^{**}$ for $ERREPQ_1$. Thus, it is chosen for visualization. The corresponding survey questions are as follows:

- EXPCURRUNIT: Years of experience in the current unit.
- ORGTRUSTQ5: I can rely on my peers/colleagues to lend me hand (help me) if I needed it.
- ORGTRUSTQ10: Most of my peers/colleagues efficiently do their work even if the unit manager is not around.
- WARMCLIMQ7N: People in this unit really do not trust each other.

SOM is built upon this top predictor variable set and the outcome variable.

In the visualization of SOMs (see Figs. 6–10), the bright orange color is associated with the higher value of the variable, while the dark blue color means a lower value of the variable. The color code is represented by the color reference bar on the right.

The number on the top of each cell indicate the SOM unit number; the bottom number is the variable value of the cell.

Each cell represents a group of subjects that are similar in the selected measurements (variables). The subjects are grouped in the same way for all the predictor variables maps and the outcome variable map.

The outline marks several region of interests on the visualization maps, that is revealing important information for nurses' willingness to report medication errors.

6.1.1. Region one: cells 1, 2, 3, 7, 8, 9, and 13

Subjects in these cells have high values (above 2.3) for the outcome variable 1, $ERREPQ_1$, which indicates that they are more willing to report when a mistake is made, but caught and corrected before affecting the patient. The outstanding characteristic for them is that they have been worked on average a very long time in the current unit: between 14 years

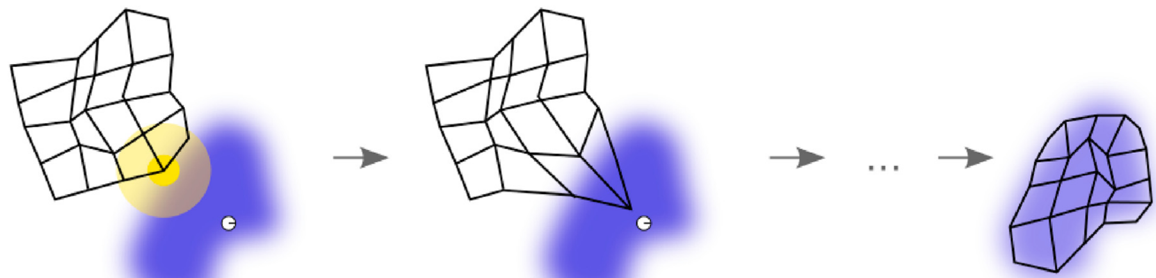

Fig. 4. An illustration of the training of a self-organizing map. The blue blob is the distribution of the training data, and the small white disc is the current training datum drawn from that distribution. At first (left) the SOM nodes are arbitrarily positioned in the data space. The node (highlighted in yellow) which is nearest to the training datum is selected. It is moved towards the training datum, as (to a lesser extent) are its neighbors on the grid. After many iterations, the grid tends to approximate the data distribution (right) [48]. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

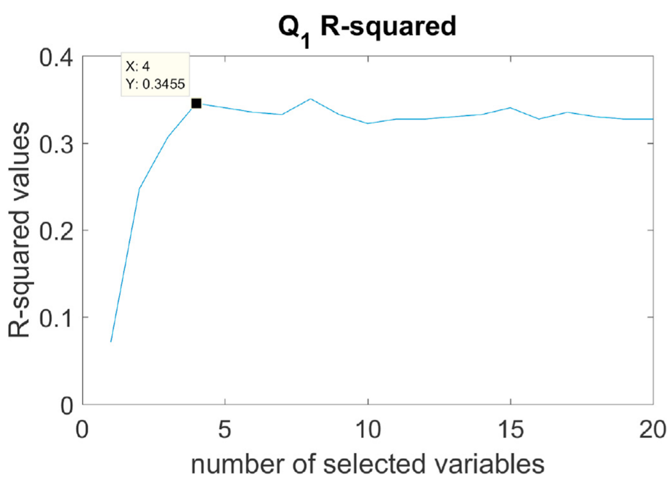

Fig. 5. R^2 Values for.$\mathbf{Y}_1$

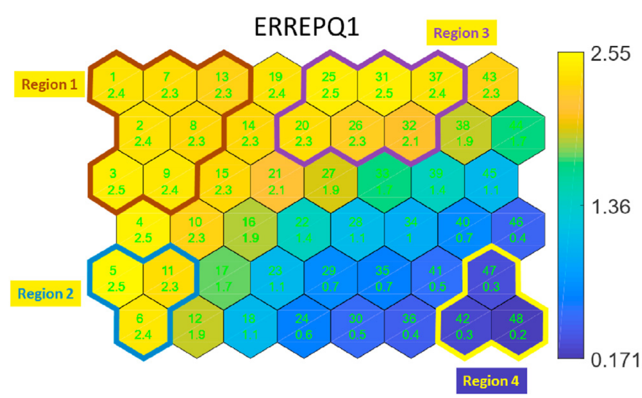

Fig. 6. When a mistake is made, but caught and corrected before affecting the patient, how likely are you to report this error? 0: Not Likely at All; 1: Somewhat Not Likely; 2: S omewhat Likely; 3: Very Likely.

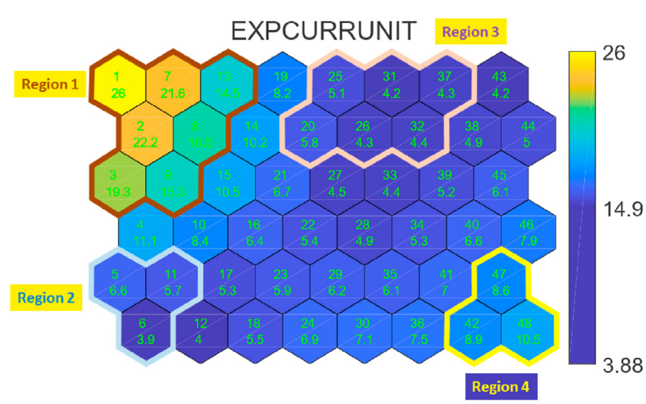

Fig. 7. How long you have been working in your current unit?.

and 26 years (indicating by the *EXPCURRUNIT* map). However, in general these subjects do not give very high score for the peer trust questions (indicating by the rest of the maps).

Conclusion: subjects have worked in the current unit for over 14 years are likely to report the $ERREPQ_1$ error.

6.1.2. Region two: cells 5, 6, and 11

Subjects in these cells also give above average scores for the variable $ERREPQ_1$. It can be noticed easily that they all worked in the current unit for 4–6 years, which is relatively short comparing to subjects in other cells. Moreover, they tend to trust their peers very much, giving very high score (around 3) to $ORGTRUSTQ_5$ and $ORGTRUSTQ_{10}$, and very low score to $WARMCLIMQ_7N$, which is a reverse question (the lower the score, the higher they feel trust).

Conclusion: subjects have worked in the current unit for under 6 years, but have very high trust levels for their peers are likely to report $ERREPQ_1$ error.

Table 4
Selected variables for ERREPQ1.

	Variable Names				
$\mathbf{S}_1^{**}$	EXPCURRUNIT				
$\mathbf{S}_2^{**}$	YSOFRNEXP	WORKHRS			
$\mathbf{S}_3^{**}$	EXPCURRUNIT	SHIFTWRK	LSHPQ20		
$\mathbf{S}_4^{**}$	EXPCURRUNIT	WARMCLIMQ7N	ORGTRUSTQ5	ORGTRUSTQ10	
$\mathbf{S}_5^{**}$	EXPCURRUNIT	SHIFTWRK	LSHPQ20	ORGTRUSTQ12N	FAMILIAREXTENTQ4

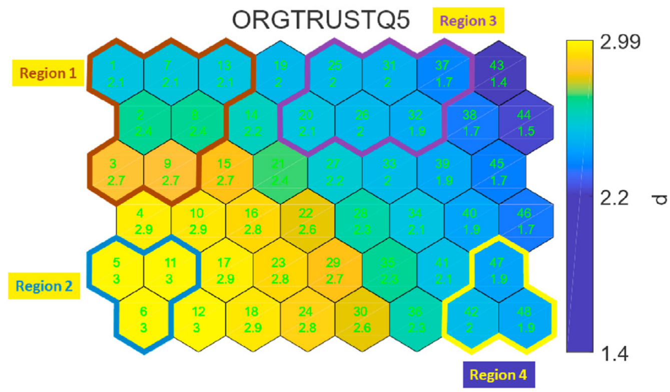

Fig. 8. I can rely on my peers/colleagues to lend me hand (help me) if I needed it. 0: Definitely Disagree; 1: Inclined to Disagree; 2: Inclined to Agree; 3: Definitely Agree.

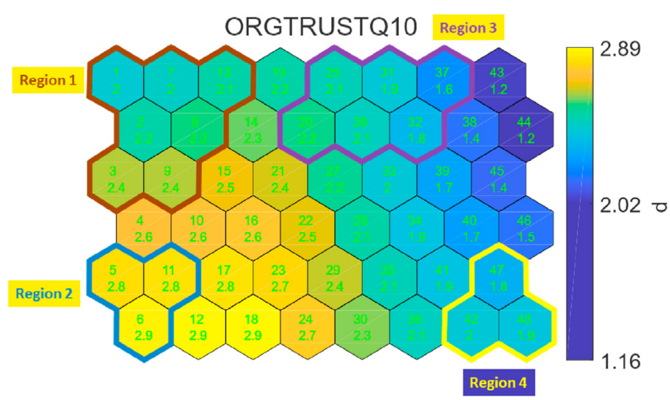

Fig. 9. Most of my peers/colleagues efficiently do their work even if the unit manager is not around. 0: Definitely Disagree; 1: Inclined to Disagree; 2: Inclined to Agree; 3: Definitely Agree.

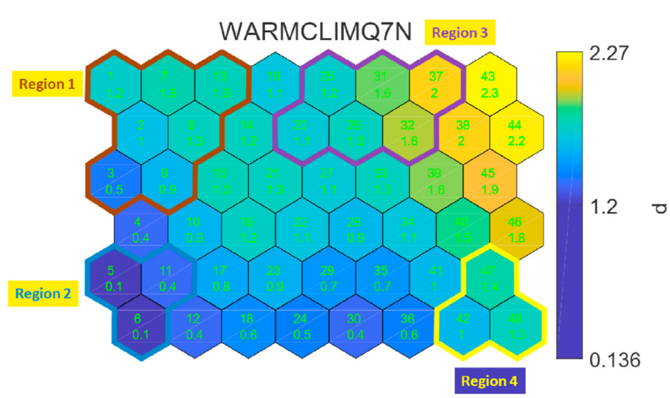

Fig. 10. People in this unit really do not trust each other. 0: Definitely Disagree; 1: Inclined to Disagree; 2: Inclined to Agree; 3: Definitely Agree.

6.1.3. Region three: cells 20, 25, 26, 31, 32 and 37

Subjects in theses cells are more willing to report as well. They are also relatively young to the current unit, between 4 and 5 years. However, their trust to the peers are not too strong, on the margin of the low trust level: around 2 for both *ORGTRUST* questions and between 1 for to 2 for the *WARMCLIM* question.

Conclusion: subjects have worked in the current unit for around 5 years, but somehow feel the lack of the peer trust, are likely to report $ERREPQ_1$ error.

6.1.4. Region four: cells 42 47 and 48

Subjects in these cells are very unwilling to report the error (average score is around or bellow 0.3). They worked in the current unit for 8–10 years. They feel somewhat trust among peers but far from strong trust.

Conclusion: subjects who have very high trust and who have very low trust are both likely to report the error. However, subjects who have medium or medium low level of peer trust are uncertain whether they will report the error or not. How long have they been working in the unit also has some effect on the subjects for reporting the error.

6.2. ERREPQ2

ERREPQ2: When a mistake is made, but has no potential harm to the patient, how likely are you to report this error?

The R^2 values for the best k variables are showing in Fig. 11.

The best variable sets $\mathbf{S}_k^{**}$, where $k = 1, 2, \ldots 20$, for $\mathbf{Y}_2$ are selected by ELMs. The following Table 5 lists selected variable names for $k = 1$ to $k = 5$.

R^2 value increases significantly at $\mathbf{S}_3^{**}$ for $ERREPQ_2$. Thus, it is chosen for visualization. The corresponding survey questions are as follows:

- SHIFTWRK: Typical working shift.
- LSHPQ5: Seeks differing perspectives when solving problems.
- LSHPQ9: Talks enthusiastically about what needs to be accomplished.

SOM is built upon this top predictor variable set and the outcome variable. The colored map from SOM are presented in Figs. 12–15.

6.2.1. Region one: cells 32, 37–40, and 43-48

Subjects in these cells are somewhat likely or very likely to report the error. The outstanding character for these subjects is that they all give very high score for the two unit manager leadership measurement questions.

Conclusion: subjects who believe their unit manager is creative when solving the problems and has enthusiasm about the goal are likely to report the error.

6.2.2. Region two: cells 6, 12, 18, 24, and 30

Subjects in these cells are not likely at all or somewhat unlikely to report the error. However the reason why they are not motivated to report is not obvious. For subjects in the cell 6 and 12, the low recognition level of the unit manager's leadership may cause the unwillingness to report. For the rest subjects the long work-shift (many of the subjects in

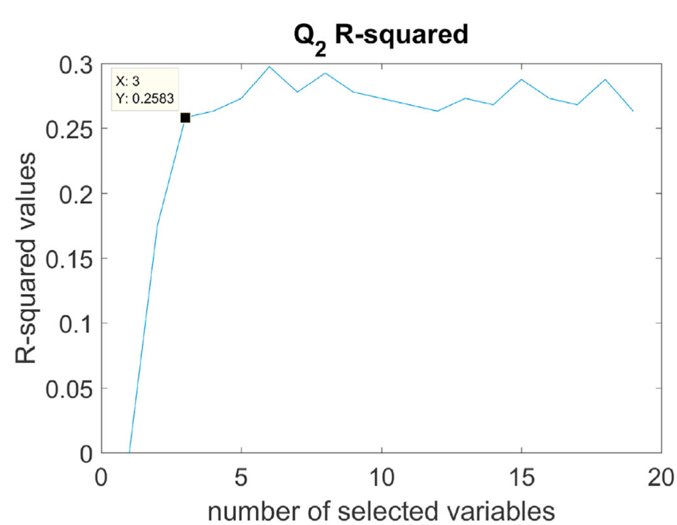

Fig. 11. R^2 Values for.$\mathbf{Y}_2$

Table 5
Selected variables for ERREPQ2.

	Variable Names				
$\mathbf{S}_1^{**}$	LSHPQ5				
$\mathbf{S}_2^{**}$	Age	WARMCLIMQ8N			
$\mathbf{S}_3^{**}$	SHIFTWRK	LSHPQ5	LSHPQ9		
$\mathbf{S}_4^{**}$	Age	WARMCLIMQ8N	ORGTRUSTQ5	ORGTRUSTQ8	
$\mathbf{S}_5^{**}$	LSHPQ1	LSHPQ10	LSHPQ11	LSHPQ17	ORGTRUSTQ5

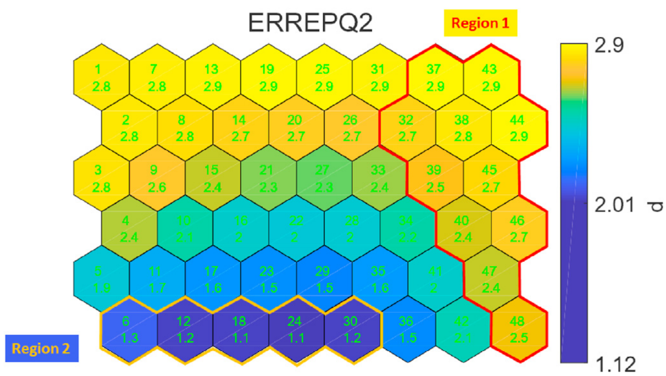

Fig. 12. When a mistake is made, but has no potential harm to the patient, how likely are you to report this error? 0: Not Likely at All; 1: Somewhat Not Likely; 2: Somewhat Likely; 3: Very Likely.

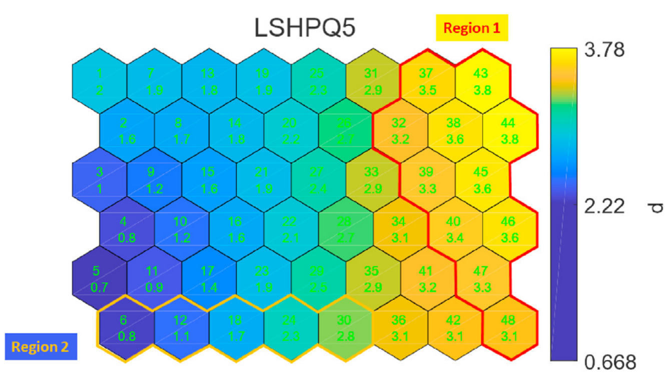

Fig. 13. Seeks differing perspectives when solving problems. 0: Not at all; 1: Once in a while; 2: Sometimes; 3: Fairly often; 4: Frequently if not always.

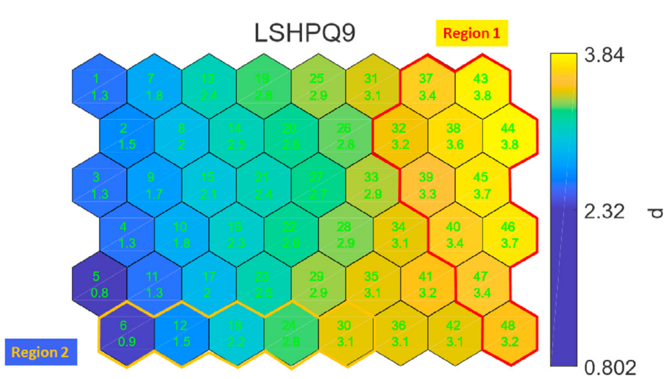

Fig. 14. Talks enthusiastically about what needs to be accomplished. 0: Not at all; 1: Once in a while; 2: Sometimes; 3: Fairly often; 4: Frequently if not always.

these cell are working at a 12 h work-shift) may be the reason of lack of motivation to report.

Conclusion: subjects who work at a long shift and think their manager are not seeking differing perspective when solving the problems or lack of enthusiasm when speaking of the goals are unlikely to report the error.

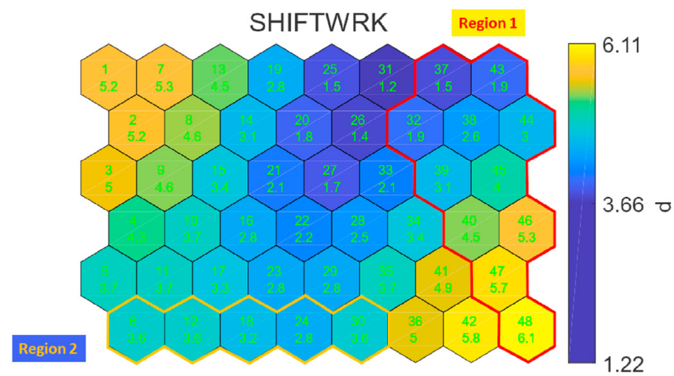

Fig. 15. Please indicate your typical shift (the shift that your work most of your time). 0 = 7am-3pm; 1 = 3pm–11pm; 2 = 11pm-7am; 3 = 7am-7pm; 4 = 7pm-7am; 5 = 8am-5pm; 6 = other; 7 = no specific shift/rotating.

6.3. *ERREPQ3*

ERREPQ3: When a mistake is made that could harm the patient, but does not, how likely are you to report this error?

The R^2 values for the best k variables are showing in Fig. 16.

The best variable sets $\mathbf{S}_k^{**}$, where $k = 1, 2, \ldots 20$, for $\mathbf{Y}_3$ are selected by ELMs. The following Table 6 lists selected variable names for $k = 1$ to $k = 5$.

R^2 value increases significantly at $\mathbf{S}_3^{**}$ for *ERREPQ*$_3$. Thus, it is chosen for visualization. The corresponding survey questions are as follows:

- EXPCURRUNIT: Years of experience in the current unit.
- LSHPQ6: Talks optimistically about the future.
- ORGTRUSTQ6: My unit manager seems to do an efficient job.

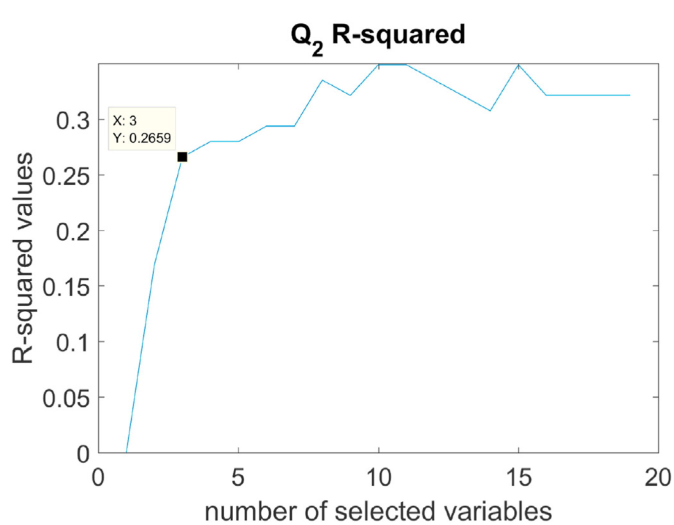

Fig. 16. R^2 Values for.$\mathbf{Y}_3$

Table 6
Selected variables for ERREPQ3.

	Variable Names				
S_1^{**}	LSHPQ18				
S_2^{**}	Age	WORKHRS			
S_3^{**}	EXPCURRUNIT	LSHPQ6	ORGTRUSTQ6		
S_4^{**}	Age	LSHPQ15	LSHPQ17	LSHPQ20	
S_5^{**}	LSHPQ10	LSHPQ18	ORGTRUSTQ6	COMPLERREPQ2	ERREPTIMEQ5

SOM is built upon this top predictor variable set and the outcome variable. The colored map from SOM are presented in Figs. 17–20.

6.3.1. Region one: cells 31, 32, 37, 38, 43 and 44

Subjects in these cells are somewhat likely to report the error. For the out come variable *ERREPQ*$_3$ the majority people are choosing very likely to report. However, for this region, subjects are hesitating. The subjects belive their manager are very optimistic about the future according to the *LSHPQ*$_6$ map, but they don't think their manager can do his/her job efficiently.

Conclusion: Subjects who have some doubts about their manager's efficiency and think the manager is optimistic about the future are somewhat likely to report the error.

7. Conclusions and future work

Results of this data analysis using SOM showed that nurses' willingness to report medication error is contingent on three factors of experience in the unit, nursing experience, organizational trust particularly trust in peers, and nurse manager leadership behaviors. Furthermore, the results showed that outcome predictors varied based on level of error severity. Based on this result, hospital administrators should consider focusing on the previously outlined predictors if they want to improve nurses' willingness to report medication errors regardless its level of severity. Using SOM, accounted for the non-liner relationship that exist among the different study variables. Most importantly it showed the pattern of organizational trust development. This information was not evident when we used traditional liner modeling.

The new methodology that is combining ELMs and SOMs has provided an clear understanding of the studied dataset. Some of the analysis are obviously right and similar to the conclusions that can be obtained with traditional data analysis. Nevertheless, more understanding has been obtained. For example, the model is sparse (few variables). It is a well-known results in the field of perception that only 5 to 6 variables can be easily understood by humans [49,50]. Furthermore unknowns nonlinear interactions between variables have been discovered using our approach. It has to be mentioned that our methodology is suitable for big

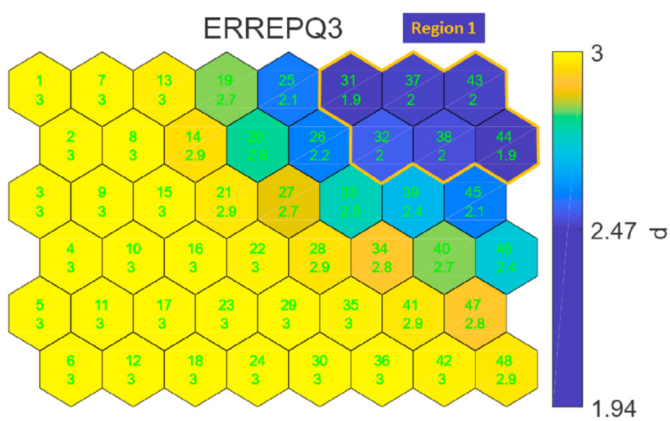

Fig. 17. When a mistake is made, that could harm the patient, but does not, how likely are you to report this error? 0: Not Likely at All; 1: Somewhat Not Likely; 2: Somewhat Likely; 3: Very Likely.

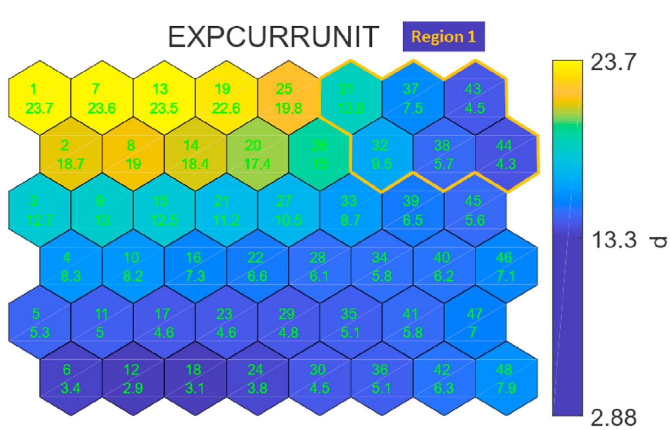

Fig. 18. How long you have been working in your current unit?.

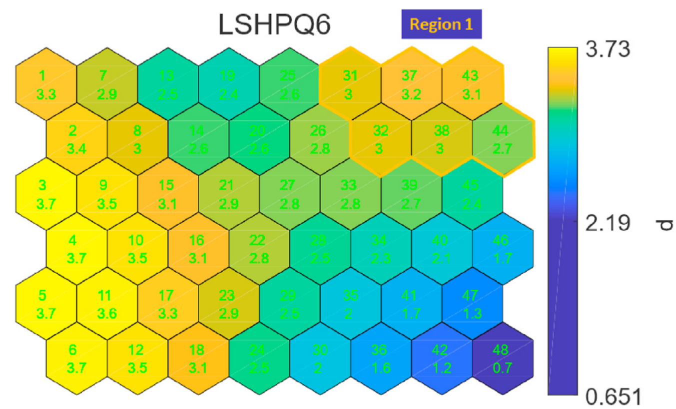

Fig. 19. Talks optimistically about the future. 0: Not at all; 1: Once in a while; 2: Sometimes; 3: Fairly often; 4: Frequently if not always.

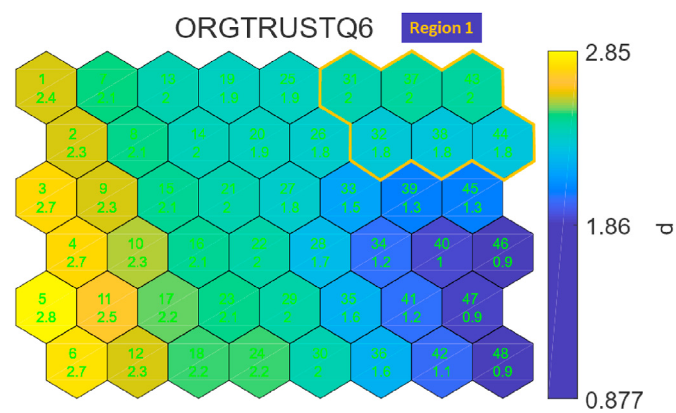

Fig. 20. My unit manager seems to do an efficient job. 0: Definitely Disagree; 1: Inclined to Disagree; 2: Inclined to Agree; 3: Definitely Agree.

data: it can handle the 3 attributes of big data: Volume, Velocity and Variety.

In the future, we are planing to use the same methodology to other medical and nursing problems. It is important to work together with practitioners to validate the results but we are willing to make the methodology nearly automatic and usable by any person that does not have a strong background in machine learning.

Author Contributions

Renjie Hu: Main contributor (all section writing, doing experiments). Amany Farag: Nursing Survey Data collecting, preliminary analysis, Introduction, data description, text checking. Kaj-Mikael: ideas and experiments and text checking. Amaury Lendasse: ideas and experiments and text checking.

Author Declaration

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from: rh2703@columbia.edu.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Institute of Medicine. To Err is human: building a safer health system. Washington, DC: The National Academies Press; 2000.
[2] M. A. Makary, M. Daniel, Medical error—the third leading cause of death in the us, BMJ 353.
[3] Janet C, Alison H, Tracey B. Nurses' role in medication safety. J Nurs Manag 2010; 18(7):853–61.
[4] Balas M, Scott L, Rogers A. Frequency and type of errors and near errors reported by critical care nurses. Can J Nurs Res 2006;38(2):24–41.
[5] Brady A, Malone A, Fleming S. A literature review of the individual and systems factors that contribute to medication errors in nursing practice. J Nurs Manag 2009; 17:679–97.
[6] Soon KK, So-Hi K, Jin-A K, Sunhee C. Nurses' perceptions of medication errors and their contributing factors in South Korea. J Nurs Manag 2011;19(3):346–53.
[7] Santell JP, Hicks RW, McMeekin J, Cousins DD. Medication errors: experience of the United States pharmacopeia (usp) medmarx reporting system. J Clin Pharmacol 2003;43(7):760–7.
[8] Reason J. Human error: models and management. Br Med J 2000;320(7237): 768–70.
[9] Mayo A, Duncan D. Nurse perceptions of medication errors: what we need to know for patient safety, vol. 19; 2004. p. 209–17.
[10] H Jones J, Treiber L. When the 5 rights go. wrong 2010;25:240–7.
[11] L. Lefrak, Moving toward safer practice: reducing medication errors in neonatal care, J Perinat Neonatal Nurs 16 (2).
[12] Benn J, Koutantji M, Wallace L, Spurgeon P, Rejman M, Healey A, Vincent C. Feedback from incident reporting: information and action to improve patient safety. BMJ Qual Saf 2009;18(1):11–21.
[13] J Wakefield B, Wakefield D, Uden-Holman T, Blegen M. Nurses' perceptions of why medication errors occur 1998;7:39–44.
[14] Cohen H, Robinson ES, Mandrack M. Getting to the root of medication errors: survey results. Nursing 2003;33(9):36–45.
[15] C. L. Covell, J. A. Ritchie, Nurses' responses to medication errors: suggestions for the development of organizational strategies to improve reporting, J Nurs Care Qual 24 (4).
[16] Performance comparison of nonlinear and linear regression algorithms coupled with different attribute selection methods for quantitative structure - retention relationships modelling in micellar liquid chromatography. J Chromatogr A 2020; 1623:461146.
[17] Tomberg A, Johansson MJ, Norrby P-O. A predictive tool for electrophilic aromatic substitutions using machine learning. J Org Chem 2019;84(8):4695–703. pMID: 30336024.
[18] Talebi M, Schuster G, Shellie RA, Szucs R, Haddad PR. Performance comparison of partial least squares-related variable selection methods for quantitative structure retention relationships modelling of retention times in reversed-phase liquid chromatography. J Chromatogr A 2015;1424:69–76.
[19] Goodarzi M, Jensen R, Vander Heyden Y. Qsrr modeling for diverse drugs using different feature selection methods coupled with linear and nonlinear regressions. J Chromatogr B 2012;910:84–94. chemometrics in Chromatography.
[20] Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: a new learning scheme of feedforward neural networks. In: Neural Networks, 2004. Proceedings. 2004 IEEE international joint conference on, vol. 2. IEEE; 2004. p. 985–90.
[21] Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: theory and applications. Neurocomputing 2006;70(1–3):489–501.
[22] Huang G-B, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification., Systems, Man, and Cybernetics, Part B: Cybernetics. IEEE Transactions on 2012;42(2):513–29.
[23] Cambria E, et al. Extreme learning machines [trends controversies]. IEEE Intell Syst 2013;28(6):30–59.
[24] Kohonen T. Self-organized formation of topologically correct feature maps. Biol Cybern 1982;43(1):59–69.
[25] Lendasse A, Cottrell M, Wertz V, Verleysen M. Prediction of electric load using kohonen maps - application to the polish electricity consumption. In: Proceedings of the 2002 American control conference (IEEE cat. No.CH37301), vol. 5; 2002. p. 3684–9. 5.
[26] Dablemont S, Simon G, Lendasse A, Ruttiens A, Blayo F, Verleysen M. Time series forecasting with SOM and local non-linear models-application to the DAX30 index prediction. In: Proceedings of the workshop on self-organizing maps, Kitakyushu, Japan, Citeseer; 2003.
[27] Merlin P, Sorjamaa A, Maillet B, Lendasse A. X-som and l-som: a double classification approach for missing value imputation. Neurocomputing 2010;73(7): 1103–8 [advances in Computational Intelligence and Learning].
[28] Lee JA, Verleysen M. Nonlinear dimensionality reduction. first ed. New York: Springer-Verlag; 2007.
[29] Kaski S, Peltonen J. Dimensionality reduction for data visualization. IEEE Signal Process Mag 2011;28(2):100–4.
[30] Lee JA, Lendasse A, Verleysen M. Nonlinear projection with curvilinear distances: isomap versus curvilinear distance analysis. Neurocomputing 2004;57:49–76. new Aspects in Neurocomputing: 10th European Symposium on Artificial Neural Networks 2002.
[31] Hu R, Ratner K, Ratner E, Miche Y, Björk K-M, Lendasse A. ELM-SOM+: a continuous mapping for visualization. Neurocomputing 2019;365:147–56.
[32] Kireeva NV, Ovchinnikova SI, Tetko IV, Asiri AM, Balakin KV, Tsivadze AY. Nonlinear dimensionality reduction for visualizing toxicity data: distance-based versus topology-based approaches. ChemMedChem 2014;9(5):1047–59.
[33] Wikipedia contributors. Feature selection — wikipedia, the free encyclopedia. available at: https://en.wikipedia.org/w/index.php?title=Feature/_selection/&oldid=849580613.
[34] Miche Y, Sorjamaa A, Bas P, Simula O, Jutten C, Lendasse A, OP-ELM. Optimally pruned extreme learning machine, Neural Networks. IEEE Transactions 2010;21(1): 158–62.
[35] A. Akusok, D. Veganzones, Y. Miche, K.-M. Björk, P. D. Jardin, E. Severin, A. Lendasse, MD-ELM: originally mislabeled samples detection using OP-ELM model, Neurocomputing.
[36] Miche Y, van Heeswijk M, Bas P, Simula O, Lendasse A. Trop-elm: a double-regularized elm using lars and tikhonov regularization. Neurocomputing 2011; 74(16):2413–21.
[37] Huang G-B, Chen L, Siew C-K. Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans Neural Network 2006;17(4):879–92.
[38] G.-B. Huang, What are Extreme learning machines? Filling the gap between frank rosenblatt's dream and john von Neumann's puzzle, [Cognitive Computation].
[39] Huang G-B. An insight into extreme learning machines: random neurons, random features and kernels. Cognitive Computation 2014;6(3):376–90.
[40] Gritsenko A, Sun Z, Baek S, Miche Y, Hu R, Lendasse A. Deformable surface registration with extreme learning machines. In: International conference on Extreme learning machine. Springer; 2017. p. 304–16.
[41] Rosenblatt F, Nonr C. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev 1958;65(6):386–408.
[42] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97.

[43] Glantz SA, Slinker BK, Neilands TB. Primer of applied regression and analysis of variance, vol. 309. New York: McGraw-Hill; 1990.
[44] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157–82.
[45] Meiri R, Zahavi J. Using simulated annealing to optimize the feature selection problem in marketing applications. Eur J Oper Res 2006;171(3):842–58.
[46] Akusok A, Baek S, Miche Y, Bjork K-M, Nian R, Lauren P, Lendasse A. ELMVIS+: fast nonlinear visualization technique based on cosine distance and extreme learning machines. Neurocomputing 2016;205:247–63.
[47] Hu R, Roshdibenam V, Johnson HJ, Eirola E, Akusok A, Miche Y, Björk K-M, Lendasse A. Elm-som: a continuous self-organizing map for visualization. In: 2018 international joint conference on neural Networks (IJCNN). IEEE; 2018. p. 1–8.
[48] Wikipedia. Self-organizing map. https://en.wikipedia.org/wiki/Self-organizing_map.
[49] Tufte ER. The visual display of quantitative information. Cheshire, CT, USA: Graphics Press; 1986.
[50] Ware C. Information visualization: perception for design. third ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2012.
[51] Halford GS, Baker R, McCredden JE, Bain JD. How many variables can humans process? Psychol Sci 2005;16(1):70–6.