

Neural Network Classifier for Automatic Detection of Invasive Versus Noninvasive Airway Management Technique Based on Respiratory Monitoring Parameters in a Pediatric Anesthesia

Jorge A. Gálvez¹ · Ali Jalali² · Luis Ahumada³ · Allan F. Simpao¹ · Mohamed A. Rehman¹

Received: 31 March 2017 / Accepted: 20 July 2017 / Published online: 23 August 2017
© Springer Science+Business Media, LLC 2017

Abstract Children undergoing general anesthesia require airway monitoring by an anesthesia provider. The airway may be supported with noninvasive devices such as face mask or invasive devices such as a laryngeal mask airway or an endotracheal tube. The physiologic data stored provides an opportunity to apply machine learning algorithms distinguish between these modes based on pattern recognition. We retrieved three data sets from patients receiving general anesthesia in 2015 with either mask, laryngeal mask airway or endotracheal tube. Patients underwent myringotomy, tonsillectomy, adenoidectomy or inguinal hernia repair procedures. We retrieved measurements for end-tidal carbon dioxide, tidal volume, and peak inspiratory pressure and calculated statistical features for each data element per patient. We applied machine learning algorithms (decision tree, support vector machine, and neural network) to classify patients into noninvasive or invasive airway device support. We identified 300 patients per group (mask, laryngeal mask airway, and endotracheal tube) for a total of 900 patients. The neural network classifier performed better than the boosted trees and support vector

machine classifiers based on the test data sets. The sensitivity, specificity, and accuracy for neural network classification are 97.5%, 96.3%, and 95.8%. In contrast, the sensitivity, specificity, and accuracy of support vector machine are 89.1%, 92.3%, and 88.3% and with the boosted tree classifier they are 93.8%, 92.1%, and 91.4%. We describe a method to automatically distinguish between noninvasive and invasive airway device support in a pediatric surgical setting based on respiratory monitoring parameters. The results show that the neural network classifier algorithm can accurately classify noninvasive and invasive airway device support.

Keywords Algorithms · Laryngeal masks · Intubation, intratracheal, masks · Neural networks (computer)

Introduction

Anesthesia provider's responsibility

Pediatric anesthesiologists must manage children's needs during surgery while documenting their physiologic state in the anesthesia record. Anesthesiologists have access to electronic health record systems, known as anesthesia information management systems (AIMS), that facilitate automatic recording of physiologic monitoring data obtained during surgical procedures [1–4]. The physiologic data stored provides an opportunity to apply machine learning algorithms to identify clinically relevant patterns that can be potentially used to provide decision support for physicians [5].

Respiratory monitoring

Children under general anesthesia may breathe spontaneously without assistance or receive mechanical ventilator support.

This article is part of the Topical Collection on *Patient Facing Systems*

✉ Jorge A. Gálvez
galvezj@email.chop.edu

¹ Section of Biomedical Informatics, Department of Anesthesiology & Critical Care Medicine, The Children's Hospital of Philadelphia, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

² Section of Biomedical Informatics, Department of Anesthesiology & Critical Care Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

³ Enterprise Analytics and Reporting, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

Patients who are breathing spontaneously generate negative intrathoracic pressure during inspiration and positive intrathoracic pressure during exhalation, whereas patients requiring mechanical respiratory support receive positive intrathoracic pressure during inspiration [6]. Positive pressure ventilation may be administered manually via squeezing the ventilation bag on the anesthesia machine or by activating the ventilator on the anesthesia machine. Anesthesia machines continuously measure respiratory parameters including end-tidal CO₂, tidal volume, peak inspiratory pressure and minute ventilation. A patient's airway may be supported using various devices that can be classified as noninvasive or invasive ventilation devices. Noninvasive techniques include spontaneous ventilation with a mask, while invasive techniques include devices such as laryngeal mask airways or endotracheal tubes. Each of these support devices and ventilation techniques produce subtle differences in the patient's ventilation parameters outlined above.

Machine learning algorithms

Current physiologic monitors and ventilators provide alarms when individual parameters cross a specific threshold, such as tidal volume below or above a certain value. This results in a variety of alerts, most of which may not indicate clinical problems but can distract the clinicians, producing alert fatigue [7, 8]. Machine learning systems that can accurately identify these changes have the potential to improve clinical monitors and documentation systems by reducing alert fatigue. Machine learning techniques have increasingly attracted attention from researchers in the biomedical field as a result of their superior performance in comparison with traditional statistical-based approaches in prediction, modeling, and classification of biomedical systems [9–11]. The main reasons for such performance include: (i) unlike the statistical approaches, the machine learning methods do not rely on any hypothesis to carry out the analysis. As a result, machine learning systems achieve unbiased results and can discover new patterns in the data [12, 13]. (ii) Logistic or linear regression methods cannot classify problems where the data is significantly complex and inherently non-linear [13, 14]. The spectrum of machine learning techniques includes data mining algorithms and methods such as decision trees, support vector machine, and neural networks. These tools facilitate data exploration using data analysis techniques with sophisticated algorithms to discover unrecognized patterns.

The goal of decision trees is to use a dataset with known attribute-class combinations for generating a tree structure with a set of rules for classification and prediction of the desired event [15]. The decision tree consists of a root, internal decision nodes and a set of terminal nodes or leaves, each representing a class. There are two phases in decision tree induction: tree building and tree pruning. In tree building phase the goal is to split the data in the way that the divided data-set is more homogenous. To this end we need to define a measure

that quantifies impurity in the data. One of the commonly used criteria is called information gain which is defined by:

$$I(t) = \sum_{i=1}^M P(w_i|t) \ln P(w_i|t) \quad (1)$$

where, t is the node, M is the number of classes which in binary classification is equal to 2, w is the class, and $|$ represents conditional probability. Boosting or boosted trees is a method of combining hundreds of smaller and weak decision trees to improve accuracy in prediction and classification, and robustness to noise in the data.

Support vector machine is a relatively new type of statistical learning machine that provides good generalization capability, which is an important prerogative in the design of any classifier [16]. Generalization capability refers to the capability of a learning machine to perform well with unseen testing data. Support vector machines aim to construct an optimal hyperplane that divide the data with minimum misclassification. The separating hyperplane is represented by equation $k(wx + b) = 0$ where, k is the nonlinear kernel, w is set of weights, x is the input vector, and b is bias. Support vector machines have been extensively used in design of clinical decision support systems [17–19].

Neural networks are a type of machine learning classifier that mimic human nervous systems through modeling neurons and their connection. The neural network consists of a series of neurons in different layers—input, hidden, and output layers—that are connected by weights. The neurons apply a function to their input which is called activation function. Multi-layer perceptron neural networks are composed of different layer. The first layer is called input layer which get the input to the network; hidden layers are then consequently connected to the input layer and map the input to the higher dimension. The final layer is the output layer which contains the probability of the output. Neural network with sufficient number of parameters and tuning is able to learn any mathematical transformation [20].

The primary hypothesis of this manuscript was to determine whether machine learning algorithms could accurately distinguish patients that have invasive ventilation with either a laryngeal mask airway or an endotracheal tube from those that have non-invasive ventilation during surgery. We evaluated the three types of machine learning algorithms: boosted tree, support vector machine and neural networks.

Methods

This study was exempt from review by the institutional review board. We retrieved de-identified electronic anesthesia records from the Clinical Data Warehouse (CDW) consisting of three groups of patients that received general anesthesia in

2015. We selected patients undergoing elective surgical procedures which to achieve a balanced group of patients with each of the ventilation devices. The inclusion criteria consisted of 1) surgical procedures: myringotomy, tonsillectomy, adenoidectomy or inguinal hernia repair, 2) documented airway device: mask, laryngeal mask airway or endotracheal tube, and 3) American Society of Anesthesiologists Physical Status Classification 1 or 2. Patients undergoing myringotomy tubes typically have natural airway with spontaneous ventilation that may transition to positive pressure ventilation. Adenoidectomy and tonsillectomy patients always have an endotracheal tube and may be breathing spontaneously or have positive pressure ventilation. Finally, patients undergoing hernia repairs may have either a laryngeal mask airway or an endotracheal tube for the procedure. Patients underwent one of the following procedures: The airway device type was determined from the clinical documentation. Respiratory parameter measurements were originally measured with the anesthesia machine (Dräger Apollo, Dräger, Lubeck, Germany) and recorded automatically in the AIMS (Epic, Verona, WI) and the clinical data warehouse in one-minute intervals (Fig. 1). The physiologic data is recorded without filtering, and zero values are included. We retrieved respiratory parameters including respiratory rate, tidal volume, end-tidal CO₂, peak inspiratory pressure from the clinical data warehouse. We calculated the minute ventilation using the following formula [21]:

$$\text{Minute Ventilation} = \text{Respiratory Rate} \times \text{Tidal Volume}.$$

Changes to physiological variables such as oxygen saturation, heart rate, or blood pressure are delayed indicator of changes in the ventilation [22]. Furthermore, changes in airway management type during routine anesthesia care does not typically have a visible effect on them. We calculated the following statistical features for each data element per patient:

mean, minimum, maximum, first quartile, third quartile, standard deviation, skewness, kurtosis, median, 10% percentile, 90% percentile, geometric mean, and harmonic mean. The geometric mean is an indicator of central tendency of the variable and is calculated using,

$$G_m(x) = \sqrt[n]{\prod_{i=1}^n x_i} \tag{2}$$

where x is the physiological measurement and n is the length of the measurement. The harmonic mean is an indicator of rates of changes in the measurement as it is an average of rates. The harmonic mean is calculated using the equation:

$$H_m(x) = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \tag{3}$$

We used the MATLAB Version R2015b (MathWorks, Natick, MA) statistical toolbox package for extracting the statistical features from the data. Each patient had a total of 65 statistical features (13 statistical features for each of the five vital signs). We reduced the size of the feature pool with the mutual information method in order to avoid over-fitting the model. The mutual information method calculates the mutual dependencies of two random variables. We ranked the features based on their mutual information with the outcome based on the following equation:

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \tag{4}$$

Where X is the feature vector, Y is the output vector, and p(x,y) is the joint probability of x and y.

The data were randomized and then divided into three sets of 300 patients with 100 in each airway device category (mask, laryngeal mask airway, and endotracheal tube) per set. The machine learning based classifier was designed with

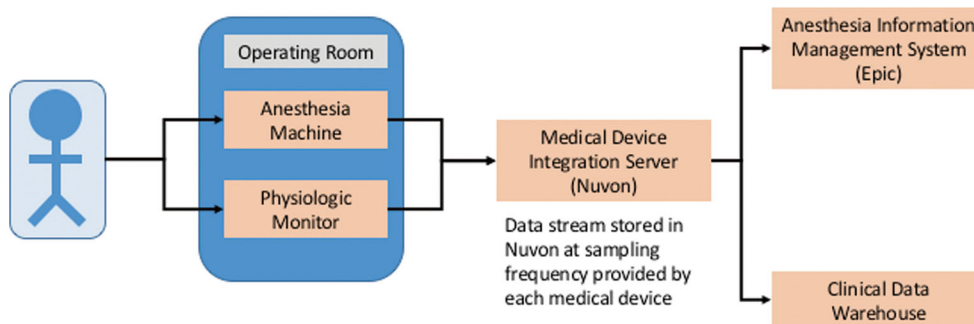


Fig. 1 Data flow from patient to the electronic health record and clinical data warehouse. The patient is attached to monitors in the operating room and is ventilated via the anesthesia machine. The physiologic data acquired through the monitors and anesthesia machine is transmitted to through the medical device integration server (MDI). The MDI server

consists of the IDM4000 gateway which transmits the data via the EGate to the electronic health record and to the Clinical Data Warehouse simultaneously. The Vega server monitors the performance of data transmission through the IDM4000 gateway

MATLAB to classify the data into two classes (Mask versus airway device: LMA, or ETT). We evaluated three different types of machine learning classifiers BTs, SVM, and NNs. We used the Mathworks® MATLAB R2015b neural networks, statistics and machine learning packages to carry on the design of the classifier and implementation of the machine learning algorithm. We designed a 5-layer feedforward neural network classifier consisting of one input layer, three hidden layers, and one output layer. The input layer was configured with 20 neurons, the hidden layers had 70, 50 and 15 neurons respectively, and the output layer had two neurons for each of the two classes (invasive and non-invasive ventilation). The input and output layer activation function is sigmoid while the hidden layers activation function is rectified linear units.

To address the bias-variance dilemma in designing the classifier, that is, making the classifier unbiased to the output by increasing the size of the neural network and suppress the resulting variance, we trained the five layers neural network using resilient back-propagation algorithm 200 times and then calculated the average for connecting weights of the neural network. We trained the neural network using the back propagation of an error. We used cross-entropy between the neural network estimation and actual output as the loss function to minimize. The cross-entropy for binary classification is defined using the following equations:

$$e_{\text{cross-entropy}} = -\frac{1}{N} \sum_{i=1}^N \left[y_i \log \hat{y}_i + (1-y_i) \log (1-\hat{y}_i) \right] \quad (5)$$

Where, $y \in \{0, 1\}$ is the actual output \hat{y} is the estimated output, and N is the batch size. The cross-entropy is proved to be a better loss function for training neural networks as its gradient is not approaching zero when the estimated error approaches zero [23]. We used L^2 regularization method to regularize the weights and biases of network to avoid overfitting [23]. We used stochastic gradient descent to optimize the loss function. The gradient descent is calculated using:

$$s' = s - \epsilon \nabla \text{loss} \quad (6)$$

Where, s' is the new estimation of parameter s that minimize the loss function and ϵ the learning rate which controls the step length at each iteration of the gradient descent algorithm. The stopping criteria for training each of the neural networks were reaching error rate of less than 0.001.

In boosted trees method we combined 200 trees. Each tree is constructed using a different subset sample from the original data. About one-third of the cases are left out of the subset sample and not used in the construction of the k th tree. For building each decision tree we used information gain criterion described by Eq. (1) to split the data. We used Fisher's exact t-test to prune redundant branches in the tree. We then used

random forest method to combine the trees. Since the random forest algorithm is built based on subsampling the data there is no need for cross-validation or a separate test set to get an unbiased estimate of the test set error.

We used soft-margin support vector machine (ν -SVM) classifier for this problem. The ν -SVM is specifically useful for the cases that high noise level will cause overlap in the data which in turn causes high number of misclassified subjects [24]. Designing the ν -SVM will lead to solve the following optimization problem:

$$\begin{aligned} \text{minimize } W(\alpha) &= \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ \text{subject to } 0 &\leq \alpha \leq \frac{1}{m} \\ \sum_{i=1}^m \alpha_i y_i &= 0 \\ \sum_{i=1}^m \alpha_i &= \nu \end{aligned} \quad (7)$$

Where, where α is the Lagrangian variable, and ν is the user-defined soft margin. We used sequential minimal optimization (SMO) which is an algorithm for solving the quadratic programming (QP) problem presented by Eq. (7).

Results

A total of 900 records were retrieved from the CDW (Table 1). We used a training set to design the classifier and then applied the test data set to determine accuracy and generalizability of the classifier. We performed the feature ranking only on the training data set to avoid any bias which may be caused by the testing data sets and increase the generalizability of the classifier's performance. The mutual information ranking showed the 20 features [end-tidal carbon dioxide mean, 2nd quantile, geometric mean, 3rd quantile, harmonic mean, 1st quantile, maximum, moment 5; peak inspiratory pressure 1st quantile, harmonic mean, geometric mean, mean; tidal volume 1st quantile, harmonic mean, geometric mean; minute ventilation harmonic mean, range, 1st quantile, 3rd quantile; respiratory rate range] that represented 90% information required for classification. Among the top 20 ranked features that were included in the classifier, eight are extracted from end-tidal carbon dioxide, four from minute ventilation, four from peak inspiratory pressure, three from tidal volume, and one from the respiratory rate. The top seven features are extracted from end-tidal carbon dioxide.

Mean represents the mean value of the recording. Generally, the mean values for mask ventilation recordings are lower than the mean values for tube and LMA because of higher amount of gas leaks in mask ventilation. Geometric mean as a feature is regularly used in situations which we want

Table 1 Composition of anesthesia records across the three data sets

		Training data set			Test data set #1			Test data set #2		
		Mask	LMA	ETT	Mask	LMA	ETT	Mask	LMA	ETT
Age (Years)	Mean	6.7	6.7	5.9	6.1	6.8	6.1	5.5	6.7	5.4
	Standard Deviation	4.3	3.7	3.9	4.9	3.9	4.1	4.8	3.2	3.8
	Minimum	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1
	Median	6	5	6	4	6	5	4	6	5
	Maximum	19	18	21	17	18	18	20	17	17
Sex	Male	54	56	71	66	62	64	61	57	77
	Female	46	44	29	34	38	36	39	43	23
ASA Status	1	38	13	60	55	11	54	52	7	58
	2	62	87	40	45	89	46	48	93	42
Length of Procedure (Minutes)	Mean	7.6	18.8	28.2	5.4	16.9	27	5.4	20.5	29.4
	Standard Deviation	6.4	15.5	15.2	3.9	8.5	14.3	4.5	20	14
	Minimum	1	0	4	1	5	0	1	5	9
	Median	5	17	24	5	15	22	4	18	26
	Maximum	34	148	85	38	62	78	29	190	102
Length of Anesthesia (Minutes)	Mean	22.1	40.2	42.2	13.9	36.2	48.6	13.9	41.2	50.6
	Standard Deviation	21	20.5	20.2	7.5	13.3	17.7	7.9	27.6	17.1
	Minimum	7	8	7	5	20	21	6	22	25
	Median	17	36.5	39	12	33	43	12	35	47
	Maximum	198	137	125	61	104	107	54	224	138

Comparison of patient characteristics and duration of procedure and anesthesia record. Each data set consists of patients with mask, laryngeal mask airway (LMA) and endotracheal tube (ETT). The American society of anesthesiologists (ASA) status is a categorical variable that is assigned by anesthesiologists caring for each patient. ASA 1 is defined by a normal healthy patient, and ASA 2 is defined as a patient with mild systemic disease. The length of procedure indicates the duration of the surgical procedure. The length of anesthesia indicates the duration of the anesthesia record, which typically begins before the procedure and ends after the procedure is completed

to compare different subjects when we have multiple recording from each subject and each of these recordings have different numeric range. Furthermore, geometric mean is often used to mitigate the effect outliers and since the data collected from intubation are very noisy due to patient movement, leaks, etc. the geometric mean can reduce the effect of noise in the data when comparing different airway management types. Harmonic mean is typically used in the cases where we want to calculate the average of rates. It is particularly useful for detection of airway management type since we are interested to understand the rate of changes in the anesthesia machine measurements. Mask ventilation has typically more sudden changes in the measurement because of gas leaks. Hence, the harmonic means of anesthesia machine measurements such as respiratory rate is usually higher for mask ventilation is comparison with endotracheal tube or laryngeal mask airway. Quartiles, including first, second, and third quartiles, and statistical moments are a way to determine the probability distribution functions of the measurements to be able to better classify them. Machine learning in general tries to learn the probability density function that generates the input to the classifier (anesthesia machine measurements) to be able to

distinguish different classes (mask vs laryngeal mask airway or endotracheal tube).

The neural network classifier performed better than boosted tree and support vector machine classifiers based on the test data sets. The sensitivity, specificity, and accuracy for each airway device neural network are 97.5%, 96.3%, and 95.8%. In contrast, the sensitivity, specificity, and accuracy are 89.1%, 92.3% and 88.3% for support vector machine and 93.8%, 92.1% and 91.4% for boosted trees respectively (Table 2). There were 10 false negative cases and nine false positive cases. We performed 5-fold cross validation on the data which means that we divided the data randomly into 5 groups of 180 patients and then trained the algorithm using 4 groups and tested it using the excluded group from training. This process was repeated 5 times and results are shown in Table 3.

The confusion rates showing the false negative, false positive, true positive and true negative rates for the neural network classifier for validation and test datasets are shown in Figs. 2 and 3. The neural network classifier performs consistently in the test data sets and works best in distinguishing patients with noninvasive ventilation via mask from those that

Table 2 Machine learning classifiers aggregate performance on test data sets based on correct identification of the airway device (mask, laryngeal mask airway or endotracheal tube)

Comparison of accuracy between the 3 machine learning classifiers.

	Neural network	Support vector machine	Boosted trees
Sensitivity	97.5%	89.1%	93.8%
Specificity	96.3%	92.3%	92.1%
Accuracy	95.8%	88.3%	91.4%

Side-by-side comparison of the classification performance between the three machine learning algorithms

have invasive ventilation with an airway device (laryngeal mask airway or endotracheal tube).

Discussion

We designed a neural network classifier that can accurately identify the use of mask ventilation versus an endotracheal or laryngeal mask airway based on physiologic data associated with the anesthesia record. These devices are commonly used caring for patients undergoing surgical procedures that require general anesthesia. Machine learning algorithms that can accurately identify clinical changes such as insertion of an endotracheal tube may be useful for various clinical applications including clinical decision support systems and information retrieval systems for electronic health records and clinical data warehouses. The machine learning algorithms we developed rely on the complete dataset and can identify transitions of airway device interventions in a dataset. This system can be applied to guide clinical documentation by accurately identifying the time that an endotracheal tube or laryngeal mask are inserted. At a minimum, these timestamps can be used as prompts to guide clinical documentation during anesthesia

Table 3 Neural network 5-fold cross validation

Group number	Sensitivity	Specificity	Accuracy
1	97.1	96.6	95.8
2	95.8	96.1	95.2
3	98.3	96.7	95.4
4	96.7	97.2	95.9
5	97.4	95.9	95.1

The neural network cross-validation consisted of randomly assigning patients to 5 groups of 180 patients. The first four groups were used to train the neural network, and the final group was used to validate it. The process was repeated five times with different groups, and the validation results for each test are presented above

care. Future applications could also include sophisticated monitor or ventilator alarms that indicate when the endotracheal tube is inserted or removed. Previous authors have demonstrated that manual documentation leads to inaccurate charting in comparison with automated documentation. This can potentially provide a more detailed description of the events in the anesthesia record, particularly in comparison with manual documentation of isolated events such as “mask ventilation.”

The reason that the neural network classifier’s performance was better than boosted tree and support vector machine classifiers could be because of its non-linear pattern mining nature. Neural network models can evaluate a large number of parameters and extract many different patterns. Meanwhile, the performance of the ensemble methods such as boosted trees performance is optimal well when the base classifiers or so-called “weak classifiers” have a weak performance. It is possible that in this case, the base classifiers have an acceptable performance, resulting in minimal gains in performance after performing the boosting procedure. Lastly, support vector machine classifiers can perform very well in binary classification problems [25]. However, the support vector machine classifiers may not demonstrate the same performance with multi-class classification.

We observed that the end-tidal carbon dioxide data was the most influential variable for the classifier’s performance. In our previous study on the detection of intubation time, we also discovered that end-tidal carbon dioxide data is the most important measurement for detection of the endotracheal intubation time [26]. One explanation is that the pattern of changes in the end-tidal carbon dioxide data is fairly independent of the mechanical ventilation mode being used to ventilate the patient. In comparison with the anesthesia machine measurements, the demographic observations such as age, gender, weight and ASA score have lower mutual information with the output of the classifier. Therefore, we did not include these variables in the training model for the classifier.

The machine learning classifier has several limitations inherent to its design. The classifiers were compared against the final clinical documentation for the airway type. It is possible that the clinical documentation was inaccurate, for example, by omission of a structured data element in lieu of text documentation. As a result, the classifier’s performance at accurately identifying the airway device could be under or over-estimated. The clinical documentation system allows providers to document the insertion of a laryngeal mask airway and endotracheal tube in structured data fields which are easily retrieved. On the other hand, mask ventilation is not documented in a structured data field. Therefore, the number of patients that had noninvasive ventilation with a mask could be under-estimated in this study in our clinical data warehouse. To address this, we retrieved anesthesia records for patients undergoing myringotomy tube insertion because the patient’s airway is managed exclusively with a mask. We

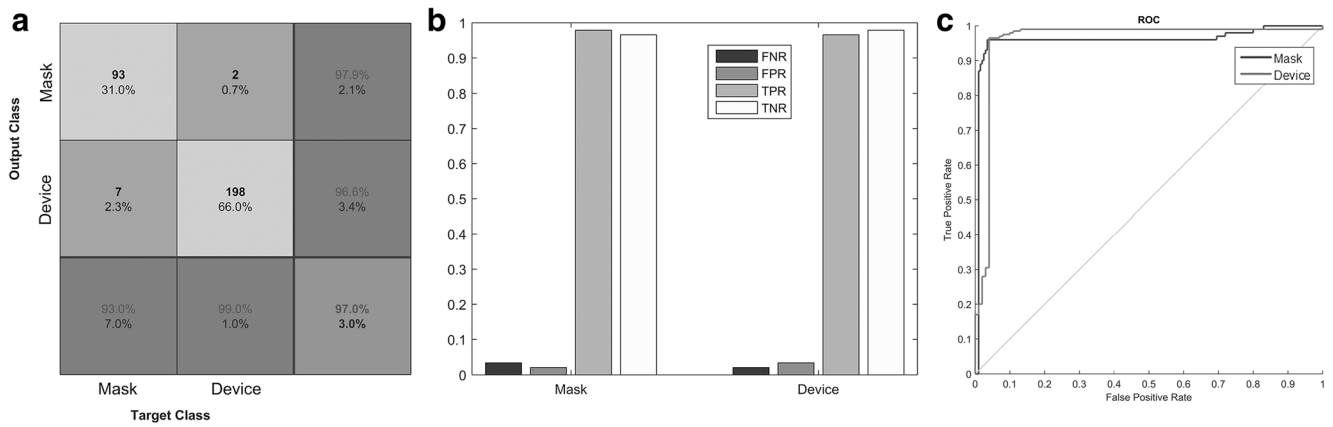


Fig. 2 Training Data Set. (Panel A) Confusion matrix for neural network classifier. (Panel B) Bar graph of false negative rate (FNR), false positive rate (FPR), true positive rate (TPR) and true negative rate (TNR) based on

the neural network classifier. (Panel C) Neural network classifier ROC graphs for the mask ventilation and airway device categories

confirmed that the procedures retrieved in the mask group did not have any documentation of a laryngeal mask airway or an endotracheal tube in the final anesthesia record. Lastly, the classifier is configured to identify only one airway device type for each procedure, which was applicable to this research study. However, in clinical practice, a patient may need multiple airway devices during a procedure. We plan to continue to develop the classifier to account for these scenarios, as well as to improve the accuracy of the detected time of transition from one airway device to another.

One potential application of this type of algorithm is in developing clinical decision support systems such as automatic documentation of timestamps for intubation and extubation. Furthermore, as these machine learning systems become more refined, they can be used to create better alarm systems to reduce alert fatigue in the operating room and intensive care units [8]. For example, the classifier could alert a clinician that the endotracheal tube has become compromised or removed in real time. This type of alert could prompt clinicians to evaluate

a patient’s status in a timely fashion. However, the algorithm needs to be validated for its performance in real time as well as identifying multiple transition points during a procedure.

Another potential application of this tool is in information retrieval for electronic health records. Anesthesia information management systems are increasingly used to create large data repositories from various institutions throughout the world [27–31]. The data quality in Anesthesia information management systems has improved over time, specifically when dealing with automatic recording of physiologic data streams [1, 2, 4, 32]. Furthermore, machine learning algorithms can assist data mining and interpretation of large and heterogeneous datasets arising from multi-institutional anesthesia information management systems. For example, identification of the airway device based on data streams that are captured automatically may retrieve records where the clinical documentation may have been incomplete. The classifier could be configured to identify records that have a potential discrepancy in airway device type documented versus what is detected. Clinicians

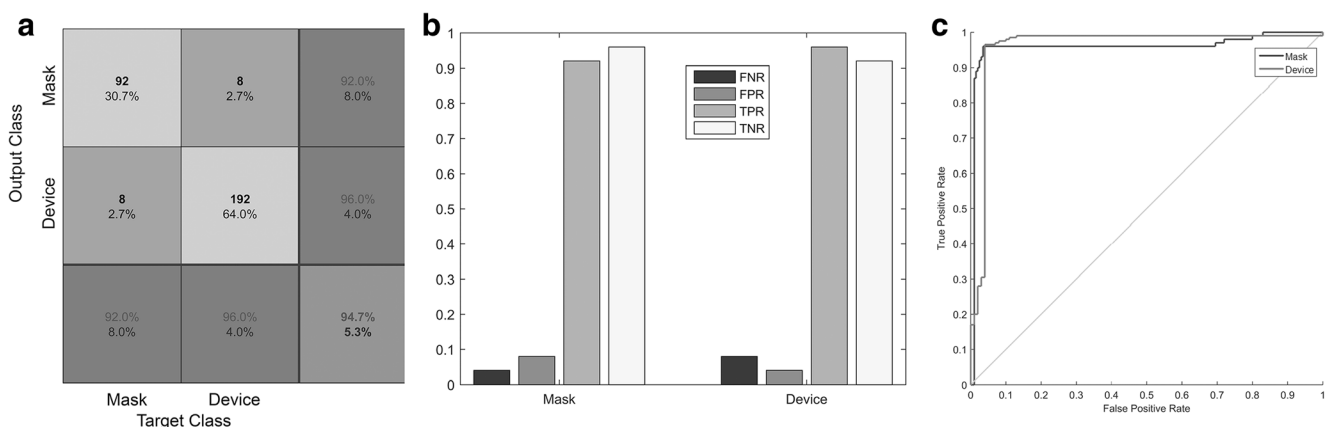


Fig. 3 Test Data Set. (Panel A) Confusion matrix for neural network classifier. (Panel B) Bar graph of false negative rate (FNR), false positive rate (FPR), true positive rate (TPR) and true negative rate (TNR) based on

the neural network classifier. (Panel C) Neural network classifier ROC graphs for the mask ventilation and airway device categories

could be notified of such discrepancies in real time to prompt a review of the documentation for accuracy and completeness.

Future work will attempt to validate the classifier's accuracy with data obtained from direct observation and classification of the airway device used. We aim to design a real-time decision support alert system that can notify health care providers about insertion and removal of invasive airway devices. These alerts can be used to alert clinicians of acute changes in a patient's status as well as to provide prompts to ensure accurate clinical documentation. Furthermore, this classifier can also be used to validate data sets for research.

Acknowledgements This work was conducted without funding support from any external agency.

Author's Contribution The following individuals contributed to 1) conception and design, acquisition of data, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published.

Jorge A. Gálvez, Luis Ahumada, Allan F. Simpao, Ali Jalali and Mohamed A. Rehman.

Jorge Galvez may be considered as a guarantor for the integrity of the work, from inception to published article.

Compliance with Ethical Standards

Conflict of Interests Jorge A. Gálvez has no conflict of interest. Ali Jalali has no conflict of interest. Luis Ahumada has no conflict of interest. Allan Simpao has no conflict of interest. Mohamed Rehman has no conflict of interest.

Ethical Approval This retrospective observational study was conducted after waiver of review was obtained from the Institutional Research Board at The Childrens' Hospital of Philadelphia because it consisted of a de-identified data set.

References

- Galvez, J.A., Rothman, B.S., Doyle, C.A., Morgan, S., Simpao, A.F., and Rehman, M.A., A narrative review of meaningful use and anesthesia information management systems. *Anesth. Analg.* 121(3):693–706, 2015.
- Reich, D.L., Wood, R.K., Mattar, R., Krol, M., Adams, D.C., Hossain, S., et al., Arterial blood pressure and heart rate discrepancies between handwritten and computerized anesthesia records. *Anesth. Analg.* 91:612–616, 2000.
- van Schalkwyk, J.M., Lowes, D., Frampton, C., and Merry, A.F., Does manual anaesthetic record capture remove clinically important data? *Br. J. Anaesth.* 107:546–552, 2011.
- Simpao, A.F., Pruitt, E.Y., Cook-Sather, S.D., Gumaney, H.G., and Rehman, M.A., The reliability of manual reporting of clinical events in an anesthesia information management system (AIMS). *J. Clin. Monit. Comput.* 26:437–439, 2012.
- Mueller, M., Wagner, C.C., Stanislaus, R., and Almeida, J.S., Machine learning to predict extubation outcome in premature infants. *Proc. Int. Jt. Conf. Neural. Netw.* 2013:1–6, 2013.
- A Practice of Anesthesia For Infants and Children. Fourth Edition ed. Charles J. Coté JL, I. David Todres, editor. Philadelphia, PA: Elsevier; 2009. 1167 p.
- Embi, P.J., and Leonard, A.C., Evaluating alert fatigue over time to EHR-based clinical trial alerts: Findings from a randomized controlled study. *J. Am. Med. Inform. Assoc.* 19:e145–e148, 2012.
- Lee, E.K., Wu, T.L., Senior, T., and Jose, J., Medical alert management: A real-time adaptive decision support tool to reduce alert fatigue. *AMIA Annu. Symp. Proc.* 2014:845–854, 2014.
- Jalali, A., Buckley, E.M., Lynch, J.M., Schwab, P.J., Licht, D.J., and Nataraj, C., Prediction of periventricular leukomalacia occurrence in neonates after heart surgery. *IEEE J. Biomed. Health Inform.* 18(4):1453–1460, 2014.
- Lehman, L., Johnson, M., Nemati, S., Adams, R., and Mark, R., *Bayesian nonparametric learning of switching dynamics in cohort physiological time series: Application in critical care patient monitoring.* *Advanced State Space Methods for Neural and Clinical Data.* 257. Harvard University OpenScholar. Cambridge University Press, Cambridge, 2015.
- Oster, J., Behar, J., Sayadi, O., Nemati, S., Johnson, A.E., and Clifford, G.D., Semisupervised ECG ventricular beat classification with novelty detection based on switching Kalman filters. *IEEE Trans. Biomed. Eng.* 62(9):2125–2134, 2015.
- Deo, R.C., Machine learning in medicine. *Circulation.* 132(20): 1920–1930, 2015.
- Haykin, S.S., *Neural networks : A comprehensive foundation.* 2nd ed. Ed. Upper Saddle River, N.J, Prentice Hall, 1999.
- Theodoridis, S., and Koutroumbas, K., *Pattern recognition*, 4th edn. Elsevier, Burlington, MA, 2008.
- Jalali, A., Licht, D.J., and Nataraj, C., Application of decision tree in the prediction of periventricular leukomalacia (PVL) occurrence in neonates after heart surgery. *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.* 2012:5931–5934, 2012.
- Cortes, C., and Vapnik, V., Support-vector networks. *Mach. Learn.* 20(3):273–297, 1995.
- Chen, M., Fu, B., Tang, T., Ma, J., and Dong, M., Hierarchical Probabilistic Support Vector Machine for Detecting Cardiovascular Diseases. *Int. J. Biosci. Biochem. Bioinforma.* 4(5):340, 2014.
- Mansson, K.N., Frick, A., Boraxbekk, C.J., Marquand, A.F., Williams, S.C., Carlbring, P., et al., Predicting long-term outcome of internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. *Transl. Psychiatry.* 5:e530, 2015.
- Zhang, Y., Lu, S., Zhou, X., Yang, M., Wu, L., Liu, B., et al., Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: Decision tree, k-nearest neighbors, and support vector machine. *Simulation.* 92(9):861–871, 2016.
- Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R., Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15(1):1929–1958, 2014.
- Barash, Paul G., Bruce F. Cullen, Robert K. Stoelting, Michael Cahalan, M. Christine Stock, Rafael Ortega, and Sam R. Sharar. *Clinical Anesthesia Fundamentals: Ebook without Multimedia.* Lippincott Williams & Wilkins, 2015.
- Ivanov, R., Weimer, J., Simpao, A.F., Rehman, M.A., and Lee, I., Prediction of critical pulmonary shunts in infants. *IEEE Trans. Control Syst. Technol.* 24(6):1936–1952, 2016.
- Goodfellow, I., Bengio, Y., Courville, A., *Deep learning:* MIT Press, Cambridge, MA, 2016.
- Scholkopf, B., Smola, A.J., Williamson, R.C., and Bartlett, P.L., New support vector algorithms. *Neural Comput.* 12(5):1207–1245, 2000.
- Scholkopf, B., and Smola, A.J., *Learning with kernels: Support vector machines, regularization, optimization, and beyond.* MIT Press, Cambridge, MA, 2001.

26. Jalali, A., Rehman, M., Lingappan, A., and Nataraj, C., Automatic detection of endotracheal intubation during the anesthesia procedure. *ASME J. Dyn. Syst. Meas. Control.* 138(11):111013-1–111013-8, 2016.
27. Abeyssekera, A., Bergman, I.J., Kluger, M.T., and Short, T.G., Drug error in anaesthetic practice: A review of 896 reports from the Australian incident monitoring study database. *Anaesthesia.* 60(3):220–227, 2005.
28. Catchpole, K., Bell, M.D., and Johnson, S., Safety in anaesthesia: A study of 12,606 reported incidents from the UK National Reporting and learning system. *Anaesthesia.* 63(4):340–346, 2008.
29. Dutton, R.P., Quality improvement and patient safety organizations in anesthesiology. *AMA J Ethics.* 17(3):248–252, 2015.
30. Dutton, R.P., Making a difference: The anesthesia quality institute. *Anesth. Analg.* 120(3):507–509, 2015.
31. Morris, G.P., and Morris, R.W., Anaesthesia and fatigue: An analysis of the first 10 years of the Australian incident monitoring study 1987-1997. *Anaesth. Intensive Care.* 28(3):300–304, 2000.
32. Benson, M., Junger, A., Michel, A., Sciuk, G., Quinzio, L., Marquardt, K., et al., Comparison of manual and automated documentation of adverse events with an anesthesia information management system (AIMS). *Stud. Health Technol. Inform.* 77:925–929, 2000.