Physician-friendly predictive model: Application to vasopressors administration prediction

Filipe Costa
filipe.n.costa@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

October 2018

Abstract

Objective: Proposal and evaluation of a predictive model for predicting vasopressors administration within one or two hours. We aim at obtaining predictive results similar or better than the results of the state of the art with the advantage that the predictions can be easily interpreted by physicians.
Materials and Methods: We considered a cohort of 25 871 intensive care unit patients, in which 22 459 did not receive vasopressors and 3 412 received vasopressors. We applied a k-Nearest Neighbor to each medical feature separately (local prediction) and we chose the most frequent prediction among the local predictions as the final prediction. Moreover, we used a feature selection process based on the characteristics of the data.
Results: Using the proposed model and the five most important features (systolic blood pressure, mean blood pressure, diastolic blood pressure, respiration rate and body temperature), we achieved AUCs of 0.908 and 0.902, for one or two hour predictions, respectively.
Discussion: The results achieved are similar to the state of the art, with the advantage that our model uses a smaller number of features and is able to provide the necessary context for physicians to accept or reject a prediction.
Conclusion: We demonstrate that it is possible to achieve good results with simple predictive algorithms and a focus on the interpretability of predictions by physicians.
Keywords: Clinical decision support, Clinical data, Predictive Models, MIMIC-III

1. Background and Significance

Over the last few years, we have witnessed a digital transformation in hospital units that resulted on the construction of huge clinical repositories containing data about past patients. Typically, these clinical repositories store data about the evolution of the values of certain clinical parameters/features (e.g., glucose, heart rate) and the treatments received during patients' stay in medical units. Therefore, these large repositories constitute an enormous potential as a source of information during the treatment of a new patient [8]. However, making decisions from the data generated in hospital units is challenging, due to the characteristics of the collected data, in particular: (*i*) *high dimensionality*: for each patient, clinical data repositories may store more than one hundred clinical features; (*ii*) *presence of outliers*: most of the data stored in clinical data repositories are introduced by humans which are error prone; (*iii*) *temporal data*: most of the features are composed by a sequence of measurements over time; and (*iv*) *unequal length temporal features*: each patient, according to her problem, may require different types of monitoring; therefore, different patients may have more measurements than others.

Constructing predictive models over data of this nature is quite challenging. There are several works that proposed predictive models to assist physicians when they are prescribing a treatment to a patient admitted into an Intensive Care Unit (ICU) [2, 9, 12, 4]. Fialho et al. [2] combine fuzzy modelling with feature selection to predict vasopressor administration within two hours in patients that required fluid resuscitation. They trained three different fuzzy models [10], one applied to the general population and the other two applied to patients with a specific disease, pneumonia or pancreatitis. The model applied to the general population achieved an Area Under the Curve (AUC)[1] of 0.79 $\pm$ 0.02; the model applied to patients with pneumonia achieved an AUC of 0.82 $\pm$ 0.02; and the model applied to patients with pancreatitis achieved an AUC of 0.83 $\pm$ 0.03. This study shows the potential of predicting vasopressors ad-

[1] AUC measures the entire two-dimensional area underneath the entire ROC curve (plot of true positive rate versus false negative rate).

ministration on the general population in ICU's and also highlights the advantages of disease specific predictive models.

The work by Salgado et al. [9] can be seen as an extension of the work by Fialho et al. [2]. They propose an ensemble fuzzy model [9] to predict vasopressor administration on patients with pancreatitis and/or pneumonia, obtaining an AUC of 0.85 $\pm$ 0.01.

Wu et al. [12] focus on three important tasks: (*i*) imminent vasopressor need (i.e., requiring vasopressor within the next two hours); (*ii*) short term vasopressor need (i.e., not requiring vasopressors for the next 4 hours but requiring it in the following 2 hours); and (*iii*) Wean readiness (i.e., a successful wean is when a patient does not require vasopressors again within 4 hours). In tasks (*i*) and (*ii*), the authors make hourly predictions until the first vasopressor administration or the end of stay. For all tasks, they use the latent states from a Switching-State Autoregressive Model (SSAM) [6] combined with raw data (physiological data + static admission data) as features, achieving respectively, for each task, AUCs of: (*i*) 0.92 $\pm$ 0.0016, (*ii*) 0.88 $\pm$ 0.0061, and (*iii*) 0.71 $\pm$ 0.005.

Later, M. Ghassemi et al. [4] tried to extend the work done by Wu et al. to the prediction of five ICU treatments: mechanical ventilation, vasopressor administration, and three blood transfusions. Again, the authors applied the latent states from a SSAM with static data of patients (e.g., gender), obtaining an AUC of 0.82 for patients that require vasopressors within the next hour.

We focus our analysis on these works [2, 9, 12, 4], because all of them used the public clinical data repository MIMIC III [5], containing health related data about 40,000 patients who stayed in critical care units of the Beth Israel Deaconess Center between 2001 and 2012. We propose to use the same data repository to enable a fair comparison.

In the previously described works we found the following limitations: (*i*) the proposed models use a large set of features (>10), which reduces the interpretability of the models. In fact, it is difficult to interpret a prediction if we need to analyze a large set of features; (*ii*) predictions are not labeled with intuitive information explaining the reasons behind them. Without this additional information, it is very difficult for physicians to understand and validate the predictions; and (*iii*) some of the proposed models are generated by complex algorithms that require time to understand and put in practice.

In this paper, we propose and evaluate a predictive model based on kNN for predicting vasopressors administration within one or two hours. The proposed model has similar results to the state of the art [2, 9, 12, 4] but with the advantage that the predictions can be easily interpreted by physicians. In particular, the proposed model overcomes the limitations of the state of the art in the following way: (*i*) it uses a small set of features (5 features) to enable physicians to easily interpret the predictions; (*ii*) it provides additional information with the predictions so that physicians can take a justified decision when accepting or rejecting the predictions. In concrete, we supply clinical information of past patients that were used for the prediction; and (*iii*) it is composed by simple steps and methods (kNN based). The proposed model achieves AUCs of 0.908 and 0.902, for one and two hours predictions, respectively.

This paper is organized as follows: Section 2 describes the materials and methods used in the construction of the proposed model; Section 3 shows the results of some experiments done in order to find the best parameterization; Section 4 compares the proposed model with the state of the art and describes the limitations and future work. Finally, Section 5 presents the conclusion.

2. Materials and Methods

2.1. Data Pre-Processing

In this study we used the MIMIC-III data repository [5] and we performed the following four data preprocessing tasks: (*i*) selection of patient records, (*ii*) identification of relevant features, (*iii*) selection of feature measurements, and (*iv*) data normalization.

The first two tasks resulted in the following patient characteristics: *Criteria 1:* patients with age $>$ 15, to exclude pediatric patients[2]; *Criteria 2:* in the case of multiple ICU's admissions, we only considered the first admission to avoid later developed complications; *Criteria 3:* patients containing at least one measurement of all the initial features considered (see Table 1); *Criteria 4:* patients who stayed in the ICU at least 24 hours after the instant where they had at least one measurement of all features (Table 1); and *Criteria 5:* patients who had at least 8 hours of clinical data before they received vasopressors. In this study, we considered the application of seven types of vasopressors: norepinephrine, epinephrine, phenylephrine, vasopressin, dopamine, dobutamine, and milrinone. A flowchart of the patient characteristics inclusion is depicted in Figure 1.

In task (*iii*), selection of feature measurements, we excluded all measurements of the initial features that were not within the acceptable interval described in Table 1. This acceptable interval per feature was defined by the team responsible for the maintenance of MIMIC. Finally, the measurement values of all features (Table 1) were normalized us-

[2] In several works that used MIMIC, patients with age higher than 15 years are seen as adults

Table 1: List of the 24 initial features considered

Feature	Unit	Category	Acceptable interval
Heart Rate	$beats/min$	Vital sign	[0, 300]
Temperature	$Celsius$	Vital sign	[0, 50]
SpO_2	$\%$	Vital sign	[0, 101]
Respiratory Rate	$breaths/min$	Vital sign	[0, 70]
Non-invasive systolic blood pressure	$mmHg$	Vital sign	[0, 400]
Non-invasive diastolic blood pressure	$mmHg$	Vital sign	[0, 300]
Non-invasive mean blood pressure	$mmHg$	Vital sign	[0, 100]
Hematocrit	$\%$	lab test	[0, 1000]
White Blood Cells	$10^3/\mu L$	lab test	[0, l000]
Platelets	$K/\mu L$	lab test	[0, l0000]
Hemoglobin	g/dL	lab test	[0, 50]
Potassium	mEq/L	lab test	[0, 30]
Sodium	mEq/L	lab test	[0, 200]
Chloride	mEq/L	lab test	[0, 10000]
Bicarbonate	mEq/L	lab test	[0, 10000]
Anion Gap	mEq/L	lab test	[0, 10000]
BUN	mg/dL	lab test	[0, 300]
Creatinine	mg/dL	lab test	[0, 150]
Glucose	mg/dL	lab test	[0, 10000]
INR	$ratio$	lab test	[0, 50]
PT	sec	lab test	[0, 150]
PTT	sec	lab test	[0, 150]
Age	$year$	static	-
Gender	$binary$	static	-

ing the *min-max* normalization technique, setting the range of all features to [0, 1].

2.2. Selection of Patient Representation

For each patient most of the features correspond to a collection of measurement values collected during an interval of time (temporal data). Temporal data is complex to handle because the amount of data is different from patient to patient and the data are collected at a wide variety of sampling frequencies. We solved the problems associated with the data with a patient representation.

In this work, we used a *closed window representation*. The time interval of measurements for each patient and feature was divided in N temporal windows of one hour. Each temporal window was represented by the average of all measurements contained in that window. The size of the window was based on the fact that the vital signs were measured every hour in the worst case. Table 2 exemplifies the representation of a generic patient with one feature (heart rate) using four windows of one hour each and considering only average as the aggregation function.

Table 2: Representation of a generic patient with one feature (heart rate), using four windows of one hour each and average as the aggregation function.

Windows	Time	Value	Window Average
window 1	00:00H 00:30H	92 94	93
window 2	01:00H 01:15H 01:30H	94 95 96	95
window 3	-	-	95
window 4	03:00H	100	100

2.3. Feature Ranking

It is important to assess the real importance of the features on the prediction of vasopressors administration in order to select a small set of features. To do that, we must answer the following question: *What are the features that better distinguish the patients who received vasopressors from the patients who did not receive vasopressors?* Two aspects need to be clarified. First, we need to define how to divide the population of patients by each feature. Second, we need to define how to evaluate the quality of the division chosen for each feature.

In clinical data we have essentially two types of features: static (e.g., age) and temporal (e.g., heart rate). Since the nature of these two types of features is different, we apply different division techniques. For static features, we apply an equal height discretization, and for temporal features, we use k-means to find groups of patients who show a similar evolution for each feature.

To evaluate the quality of the division chosen for each feature we use the *information gain ratio* [7]. For each feature we need to find the number of groups that maximizes the *information gain ratio*. To do this, we successively consider larger groups. At the end, we choose the highest value obtained as the importance of the feature.

2.4. Predictive Model

A fairly simple predictive algorithm that can be easily interpreted by physicians is k-Nearest Neighbor (kNN) [1]. This method has the advantage of exploring the natural intuition of similarity between patients. For instance, it would be interesting to

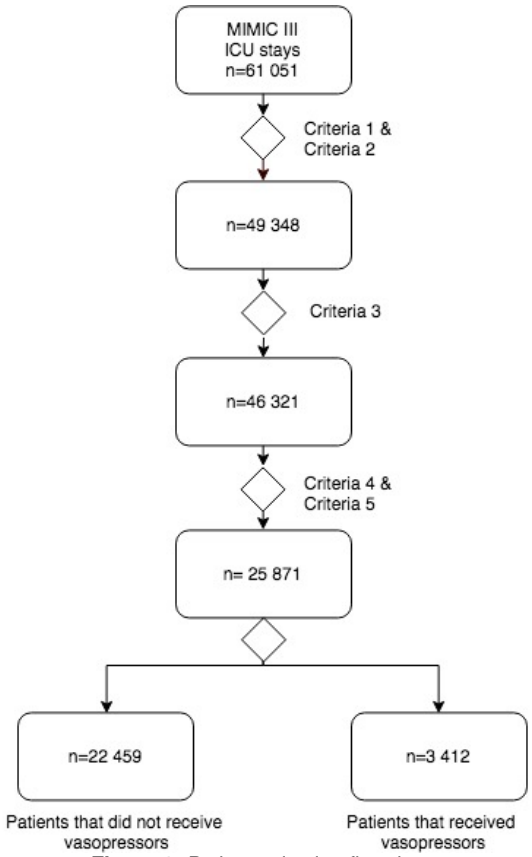


Figure 1: Patient selection flowchart

inform a physician that a patient p needs to receive vasopressors because she has a similar evolution to a patient z who received the treatment. Moreover, we can increase the interpretability of the kNN results by applying the algorithm on each feature separately (local prediction) and considering k=1. Then, the final prediction is based on the local predictions.

There are different ways of combining the local predictions to produce a final prediction. One of the simplest techniques is to choose the most frequent prediction among the locals predictions as the final prediction.

We can use the results of the feature ranking in two ways: (i) instead of the final prediction be just the most frequent prediction among the local predictions, we can weight the local predictions by their importance (normalized); and (ii) perform a feature selection process based on the importance of features. We start with the most important feature and we add in each step (to the set of features considered) the next feature more important according to the feature ranking.

2.5. Experimental Setup

Our experimental population consisted on 22 459 patients who did not receive vasopressors and 3 412 patients who received vasopressors. For patients who received vasopressors, we extracted the 7 windows (for one hour predictions) or the 6 windows (for two hours predictions) immediately before the window where patients received vasopressors. For patients who did not receive vasopressors, we randomly extracted 7 or 6 consecutive windows of the first stay.

To evaluate the performance of a model we repeated the following process ten times: (*a*) construction of a dataset where the number of patients of both types (patient who received vasopressors and patients who did not receive vasopressors) was the same; and (*a*) application of 10 fold cross validation. By repeating this process 10 times for each experiment, we obtained 100 results (10 datasets * 10 fold cross validation). At the end, we calculated the average sensitivity, specificity and AUC considering all results.

3. Results

3.1. Feature Ranking

For patients of both types (who received vasopressors and who did not receive vasopressors), we analyzed the importance of the features only considering the data before the patients received vasopressors. We considered 8 windows out of the N windows that represented each patient. For patients who received vasopressors, we considered the 8 windows before the patients received vasopressors (including the window where the patient received vasopressors). For patients who did not receive vasopressors, we randomly extracted 8 consecutive windows. By doing this, we had the same amount of information for both types of patients.

We started by identifying, for each feature, the number of groups of patients that maximizes the *information gain ratio*. Table 3 shows the features with higher values of the *information gain ratio* and the number of patient groups that maximized the value, considering eight hours of data before the patients received vasopressors.

Observing Table 3, we conclude that the systolic blood pressure (SysBP), the mean blood pressure (MeanBP) and the diastolic blood pressure (DiasBP) are the features with the highest *information gain ratio*. The importance revealed by these features is aligned with the main goal of vasopressors administration, which is to increase the arterial blood pressure. With a simple procedure based on clustering and the *information gain ratio*, we were capable of finding a feature ranking that seems to be related with the administration of vasopressors.

Table 3: Information gain ratio (importance) of features with higher value considering 8 hours of data.

Feature	Importance	Groups (K)
SysBP	0,0212	3
MeanBP	0,0115	3
DiasBP	0,0109	2
RespRate	0,0058	3
TempC	0,0057	3
PTT	0,0045	4
WBC	0,0044	3
HeartRate	0,0034	3
BUN	0,0029	2
Potassium	0,0023	4
SpO2	0,0021	4
Anion Gap	0,0021	2
Bicarbonate	0,0020	4
Creatinine	0,0014	4
Hemoglobin	0,0013	3
Platelet	0,0013	3
Sodium	0,0011	2

3.2. Predicting Vasopressors Administration

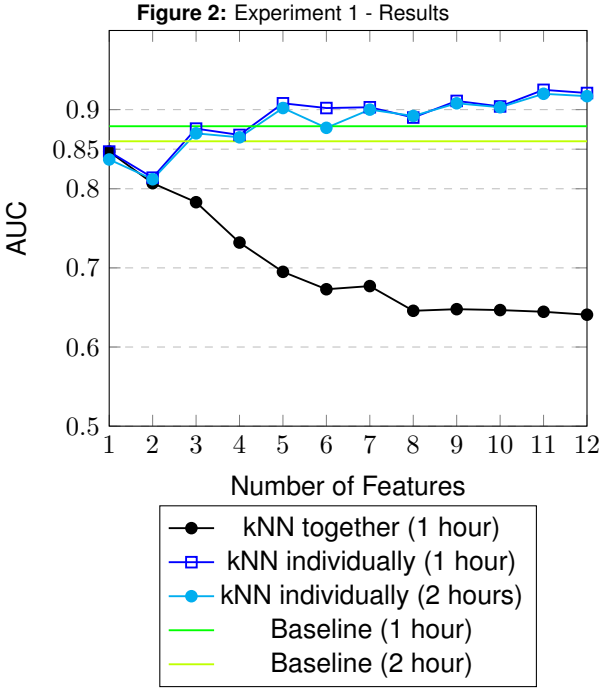
As baseline, we applied a neural network[3] to all features together (the initial 24 features). This is a well known method that typically presents high accuracy. However, this method is a "black box" with almost zero interpretability. Applying a neural network to all features together resulted in an AUC of 0.879 and 0.860, for one and two hours predictions, respectively. These were very positive results to start and show evidence that it is possible to achieve good results in the prediction of vasopressors administration.

To create a model based on kNN, several aspects need to be tested in order to find the best parameterization. We decided to perform two experiments with different goals: *Experiment 1* - evaluation of the impact of applying kNN to all features together versus applying kNN to each feature individually (the most frequent prediction among the local predictions is the final prediction and following a feature selection based on the feature ranking); and *Experiment 2* - evaluation of the impact of weighting the local predictions by the importance of the features. Figure 2 and Figure 3 show the results of these experiments.

Observing Figure 2 we conclude that: (i) the application of kNN to each feature individually outperformed kNN applied to all features together; (ii) the results of making a prediction within one hour were very similar to the results of making a prediction within two hours (iv) using only 5 features (systolic blood pressure, mean blood pressure, diastolic blood pressure, respiration rate and body temperature) on a kNN applied to each of these features individually outperformed our baseline (neural network) achieving an AUC of 0.908 and 0.902 for one and two hours predictions respectively.

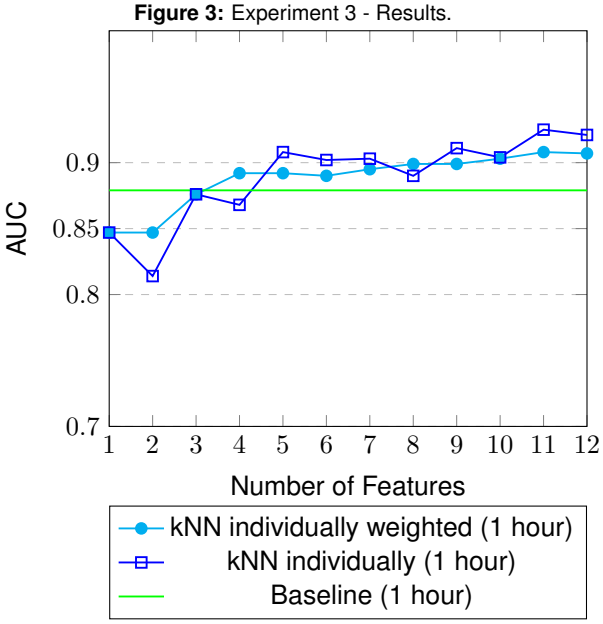
[3]Parameterization: single hidden layer with 5 Neurons, learning algorithm - BFGS algorithm [3], iterations - 200, decay - 5e-4, rang - 0.1, activation fun - logistic, error - least squares, R package - nnet [11].

Figure 2: Experiment 1 - Results



By observing Figure 3, we can see that weighting the local predictions only had a positive impact considering a small set of features.

Figure 3: Experiment 3 - Results.



4. Discussion

The results of the experiments reported in Section 3 allow us to find the best parameterization for a model based on kNN. For the prediction of vasopressors administration within one or two hours, the best model is kNN applied to each feature separately (local prediction) and the most frequent pre-

Table 4: Comparison of the results achieved (considering the feature selection based on the feature ranking and the forward feature selection) with the state of the art.

Paper	AUC	Prediction	Features
Ghassemi et al. [4]	0.820±?	1 hour	29
Proposed model	**0.908±0.01**	**1 hours**	**5**
Fialho et al. [2]	0.790±0.02	2 hours	10
Salgado et al. [9]	0.850±0.01	2 hours	24
Wu et al. [12]	0.920±0.0016	2 hours	19
Proposed model	**0.902±0.01**	**2 hours**	**5**

diction among the local predictions as the final prediction. The proposed model must use a small set of features in order to be possible for physicians to interpret the predictions. With that in mind, the best trade-off between the number of features used and the resulted AUC was achieved when we considered 5 features (systolic blood pressure, mean blood pressure, diastolic blood pressure, respiration rate and body temperature).

The comparison of our results with the state of the art [2, 9, 12, 4] is not straightforward because each work used different filtration criteria, different evaluation procedures and even different versions of MIMIC. Table 4 summarizes all the results obtained. We can see that our results are similar to the state of the art, with the advantage that our model uses a smaller number of features and is able to provide the necessary context for physicians to accept or reject a prediction. For instance using the proposed model if we want to predict if a patient p is going to need to receive vasopressors in the next hour, we just need to find the patient with the most similar evolution of the systolic blood pressure, mean blood pressure, diastolic blood pressure, respiration rate and body temperature. Imagine that for systolic blood pressure, mean blood pressure and diastolic blood pressure, a patient x is the patient who shows the most similar evolution and patient x received vasopressors. For respiration rate the most similar patient is a patient y and the patient y received vasopressors. For body temperature the most similar patient is a patient z and the patient z did not receive vasopressors. In this scenario, the proposed model will predict that patient p will receive vasopressors in the next hour. Associated with the prediction, we can provide to the physician the following information: (i) patient x evolution of systolic blood pressure, mean blood pressure and diastolic blood pressure, and say that this patient received vasopressors; (ii) patient y evolution of respiration rate, and say that this patient received vasopressors; and (iii) alert that patient p had a similar evolution of body temperature to a patient z that did not receive vasopressors. Figure 4 shows the information provided to physicians.

4.1. Limitations and Future Work

There are many interesting paths that can be further explored:

- Find the best way to communicate the predictions and respective context to the physicians;
- Assess the real impact of the window size on the patient representation and the impact of the aggregation functions used to represent the windows;
- Evaluate the impact of using other clustering algorithms (e.g. hierarchical) on the feature ranking step.

This work has two main limitations. First, we only used one clinical repository in the validation of our model. Second, static data (e.g. gender) is not fully explored by kNN applied to each feature separately.

5. Conclusion

There are different works [2, 9, 12, 4] proposing predictive models for assisting physicians when they making decisions. However, the proposed predictive models do not enable physicians to interpret the predictions. This limitation was our motivation to propose a predictive model based on kNN for predicting vasopressors administration within one or two hours. The proposed model uses only 5 features and provides additional information to enable physicians to accept or reject a prediction. Despite the simplicity of the model, we achieve AUCs of 0.908 and 0.902, for one and two hours predictions, respectively. The results achieved are aligned with the state of the art. In summary, we demonstrated that is possible to achieve good results with models based on simple algorithms and focused on the interpretability of predictions by physicians.

References

[1] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

[2] A. S. Fialho, L. A. Celi, F. Cismondi, S. M. Vieira, S. R. Reti, J. M. Sousa, and S. N. Finkelstein. Disease-based modeling to predict fluid response in intensive care units. *Methods of Information in Medicine*, 52(6):494–502, 2013.

[3] R. Fletcher. Practical Methods of Optimization. *John Wiley & Sons*, 53:456, 2013.

[4] M. Ghassemi, M. Wu, M. C. Hughes, P. Szolovits, and F. Doshi-Velez. Predicting intervention onset in the ICU with switching state space models. *AMIA Joint Summits on Translational Science proceedings.*

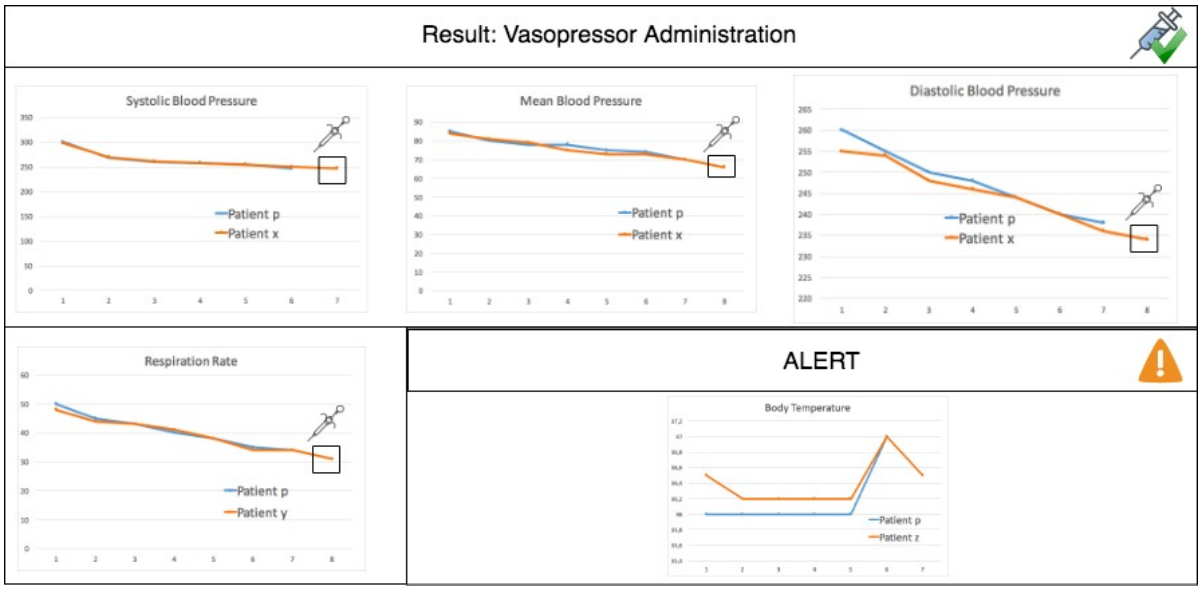


Figure 4: Example of context provided to physicians.

AMIA Joint Summits on Translational Science, 2017:82–91, 2017.

[5] A. E. W. Johnson, D. J. Stone, L. A. Celi, and T. J. Pollard. The MIMIC Code Repository : enabling reproducibility in critical care research. *Journal of the American Medical Informatics Association : JAMIA*, 25(September 2017):32–39, 2018.

[6] L. W. H. Lehman, R. P. Adams, L. Mayaud, G. B. Moody, A. Malhotra, R. G. Mark, and S. Nemati. A physiological time series dynamics-based approach to patient monitoring and outcome prediction. *IEEE Journal of Biomedical and Health Informatics*, 19(3):1068–1076, 2015.

[7] J. R. Quinlan. *C4.5: Programs for Machine Learning*, volume 1. San Mateo California, 1992.

[8] W. Raghupathi and V. Raghupathi. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1):10, 2014.

[9] C. M. Salgado, S. M. Vieira, L. F. Mendonça, S. Finkelstein, and J. M. Sousa. Ensemble fuzzy models in personalized medicine: Application to vasopressors administration. *Engineering Applications of Artificial Intelligence*, 49:141–148, 2016.

[10] T. Takagi and M. Sugeno. Fuzzy identification of system and its applications to modeling and control. In *IEEE Transactions on Systems, Man, and Cybernetics*, volume SMC-15, pages 116–132, 1985.

[11] W. N. Venables and B. D. Ripley. Statistics Complements to Modern Applied Statistics with S. In *Modern Applied Statistics with S*, page 48. Springer, 2002.

[12] M. Wu, M. Ghassemi, M. Feng, L. A. Celi, P. Szolovits, and F. Doshi-Velez. Understanding vasopressor intervention and weaning: Risk prediction in a public heterogeneous clinical time series database. *Journal of the American Medical Informatics Association*, 24(3):488–495, 2017.