2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Intelligent patient monitoring for proactive alerting of key personnel in intensive care: A single-center study

Vikas Rana[1], Teddy Le Nguyen[1], Veera Raghava[1], Prahlad G Menon[1,2]

Abstract*— A code blue event is an emergency code to indicate when a patient goes into cardiac arrest and needs resuscitation. In this paper, we model the binary response of a intensive care unit (ICU) patients experiencing a code-blue event, starting with vital time-series data of patients in 12 ICU beds. Our study introduces day-of and day-ahead risk scoring models trained against ground truth information on per-patient-per-day code-blue events, starting with multi-variate vital-time-series-sequences of varying durations with a plurality of engineered features capturing temporal variations of these signals. Actionable events, including code-blue events, aggregated by patient by day were predicted on the day-of or day-ahead with an overall accuracy of over 80% in our best models. Such models have potential to improve healthcare delivery by providing just-in-time alerting, enabling proactive and preventative clinical interventions, through continuous patient monitoring.

BACKGROUND

The rising hospital utilization rates, in ICUs and ERs around the country, has brought about a need for clinicians to develop statistical awareness and thinking as yet another critical judgment skill they bring to their patients' bedsides. Intelligent monitoring and control (IMC) can help with the interpretation of observed patient behaviors based on monitored physiological / vital signs and to plan for reasoned responses to observed events.

Almost half (48%) of hospital deaths occur in unmonitored patients [1] of which more than one-third are over 75 years of age [2]. Smart Electronic Medical Record (EMR) systems, precursors to AI-driven models, have been shown to reduce the risk of patient morbidity and mortality up to 50% [3] with studies on tele-ICUs reporting up to a 26% reduction in ICU mortality [4]. Thus, real-time AI systems are poised to profoundly impact patient care through continuous patient monitoring and risk stratification. While we aren't the first to attempt intelligent monitoring of patients [3-9] our use of high-performing AI models trained on our novel outcomes-associated vital time-series database will enable predictive monitoring of patient health.

The cost of ICU care is estimated at USD 26.2 billion a year [1]. With health expenditure escalating rapidly in the US, it is quintessential to be able to estimate ICU utilization in order to optimize management of care. There is a need for predictive models that can be utilized to exercise optimal staffing strategies and to estimate the impact of treatment approaches and patient care management paradigms that correlate with shortened length of stay (LoS)[2].

METHODS

Data Acquisition

Anonymized patient-specific vital time-series data from patients in 12 ICU beds at The Brooklyn Hospital (TBH) were acquired by leveraging HL7 feeds from a healthcare data interchange service (see Figure 1), under IRB approval, anonymizing patient identifiers at source. The vital time-series data were recorded in a manner analogous to the MIMIC-III dataset [10]. The universe of sensors in hospital beds included BluPRO P225F SpO2 sensors, Cap-ONE TG-920P CO2 Sensors, iNIBP blood pressure devices, and CardioFax-G Electrocardiographs (ECG-2550), which record of heart rate, blood pressure, respiratory rate, and oxygen saturation on a continuous basis.

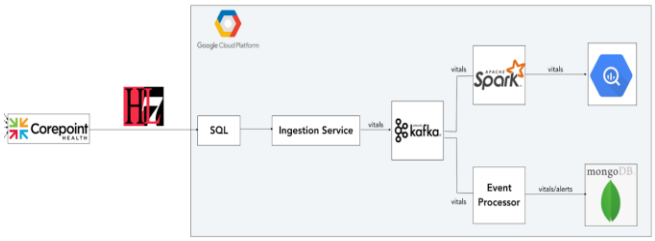

Figure 1. The architecture diagram of how we collect vitals data from sensors.

Table 1. Descriptive statistics on our dataset, stratified by period of data acquisition.

Month	Total records	# of pts	Average # obs per pt.	Total # of events	Avg daily # of events
Mar-July	2,190,077	1,407	1,557	4421	29
Aug-Sep	2,014,738	961	2,096	1663	27
Oct-Dec	2,855,911	1,345	2,123	1464	24
Jan-Feb	1,597,978	837	1909	1363	24

Lyniate Corepoint, a modular integration engine, is used to collate every patient's time-series data acquired from the

[1] Techolution Consulting LLC; [2]Dept of Bioengineering, University of Pittsburgh (email: prm44@pitt.edu);

vital sensors as HL7 streams and insert these records into a SQL database. An ingestion service streams new records from SQL into a Kafka topic on Google Cloud, every 2 minutes. A Spark streaming application listening to this topic cleanses this data before saving into a table on the BigQuery data lake service on Google Cloud. Data is fetched from source in a batch every ~142 secs whereas each observation of vitals in data in our time-series were available at a sampling rate of ~1 minute. The size of our data set in terms of records by time-range includes number of patients, vital observations over time, and the average count of observations per patient, aggregated across different time periods of data acquisition are shown in Table 1.

Model-Ready Dataset Preparation and Ground Truth: Labeled ground-truth/outcome data regarding high-risk events that occurred in the ICU, including code-blue events, was acquired monthly based on patients that visited each of TBH's 12 ICU beds. These ground truth events were time-stamped precisely only to the day of the actual event. We left-joined patient-specific ground truth back to vital time-series tables in our BigQuery data lake, based on the date of the recordings and the anonymized patient identifiers. We then split the data into timeframes based on the month of the record i.e. March-July, Aug-Sept and Oct-Dec, etc. The data collected from March-July was used to train our model while data from Aug-Sept was used for validation purposes. Data from Oct 2020 was considered as out-of-sample testing data.

Actionable Clinical Event Modeling: We define an actionable event as any event where a clinician responds to an urgent patient need; as such, this included code-blue events but other events that prompted urgent clinical attention, as well. We developed a universe of models to score individual patient risk and report early-warnings or instantaneous alerts regarding their state of physical health. As such, the goals of each model were to predict the binary response of a given patient experiencing a code-blue event. The latter was modeled on an instantaneous level and aggregated on the patient-day level and the patient-level for the purpose of reporting alerts as well as for evaluating model performance.

We developed several iterations of predictive code-blue response models, all of which were trained with Mar-July vital-time-series data with matched ground-truth events, and validated and tested with the remaining data, as described previously. A plurality of model types were explored:

a) Cross-sectional model: Instantaneous (day-of) risk scoring model were trained on per-patient-per-day code-blue information, with patient-day-level aggregation for reporting. This model had no bearing for the trends in the vital signatures but merely relied on the instantaneous observed values of a plurality of vitals and related these data to the response of code-blue events observed.

b) Two types of vital-time-series-sequence models were built for different sequence lengths (10, 50 and 100 observations per sequence), aggregated features of these time-series sequences and their first order derivatives: i) Day-of actionable event risk scoring model; and ii) Day-ahead and day-of event risk scoring model

Table 2 and 3 provide list each of the models developed and evaluated in this study.

Table 2. Day-of code-blue risk models.

Model	Method
Distributed Random Forest(DRF)	Cross-sectional model with time-series aggregation to the patient-day level for performance reporting
XGboost(XGB)	XGboost machine learning model with aggregation to the patient-day level for performance reporting
Gradient Boosting (GB)	Gradient Boosting machine learning model using sequence-level features, with sequence length=50
Deep neural network (DNN)	Deep neural network with sequence length=100
XGB-1	XGB machine learning model with Sequence length=10
XGB-2	XGB machine learning model with sequence length=10 and has features after feature selection approach

Table 3. Day-ahead and day-of code-blue risk model

Model	Method
LSTM	LSTM deep learning model with 5 layers that included LSTM, drop-out, and dense layers, built using tensorflow/keras with 100 nodes in the first LSTM layer and 50 nodes in a second LSTM layer.

Our day-of as well as day-ahead temporal sequence models for actionable event risk scoring contained statistically aggregated features (eg: min, max, mean, median, first & last value, mode, skew, kurtosis, 25th & 75th percentile, standard deviation and variance) of time series sequences from original vitals and their first order derivatives.

Model Optimization: For instantaneous vital signal level, we optimized the threshold using receiver operator characteristics (ROC) analysis. The ROC curve was created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The ROC curve helped identify the threshold where the TPR is high and FPR is low i.e. misclassifications are low. There we found the optimal probability threshold where we got the highest sensitivity and specificity figures. For patient-day level models, we applied the same ROC analysis approach; however, in addition to optimal threshold, we calculated the optimal count of positives for patient-per day aggregation. The optimal count of positives is defined as the point at which balanced sensitivity and specificity was maximized. In combining these two approaches, we identified the best optimal threshold and optimal count of positives at which we got best sensitivity and specificity figures.

RESULTS

Tables 4 through 6 summarize our model performance metrics, for each model and each dataset group utilized for training, validation and testing.

Table 4. Sensitivity, specificity, precision of per-patient-per-day code-blue risk prediction, for each model in our study, for each data subset. DRF, XGB, GB, DNN, XGB-1 and XGB-2 predict day-of events, whereas LSTM predicts day-ahead and day-of events.

ID	Metric	03/20-07/20	08/20-09/20	10/20-12/20	12/20-01/21	02/21
DRF	Sensitivity	83.09	34.50	49.67	39.42	45.00
	Specificity	83.23	73.29	60.57	64.36	59.45
	Precision	34.90	8.56	6.12	7.87	7.12
XGB	Sensitivity	60.24	64.50	61.58	50.00	61.67
	Specificity	60.65	56.23	56.55	61.09	53.11
	Precision	14.20	9.65	6.84	9.02	8.33
GB	Sensitivity	71.24	48.22	64.79	66.25	74.00
	Specificity	67.54	57.38	50.70	32.21	30.49
	Precision	19.64	7.85	6.61	7.17	7.02
DNN	Sensitivity	64.04	54.79	66.18	47.61	46.15
	Specificity	64.09	52.41	50.20	52.00	52.32
	Precision	16.85	8.00	6.84	6.68	5.71
XGB-1	Sensitivity	70.93	63.82	71.52	43.75	55.36
	Specificity	70.30	47.78	51.10	65.89	63.95
	Precision	20.45	8.19	7.19	8.99	9.45
XGB-2	Sensitivity	73.43	59.79	67.55	40.62	48.21
	Specificity	73.37	54.79	57.63	70.70	68.93
	Precision	22.89	8.81	7.79	9.65	9.54
LSTM	Sensitivity	66.24	67.35	67.21	54.07	71.60
	Specificity	65.63	33.36	37.02	56.50	36.17
	Precision	28.65	11.74	8.73	12.20	10.21

Table 5. Sensitivity, specificity, precision for patient level code-blue prediction.

ID	Metric	03/20-07/20	08/20-09/20	10/20-12/20	12/20-01/21	02/21
DRF	Sensitivity	90.13	50.00	76.51	54.94	75.92
	Specificity	92.97	72.62	60.73	67.69	61.92
	Precision	86.70	35.71	28.29	25.38	31.06
XGB	Sensitivity	74.34	78.12	78.03	62.63	81.48
	Specificity	60.86	57.03	57.05	65.49	60.66
	Precision	49.13	35.61	26.89	26.63	31.88
GB	Sensitivity	80.50	65.40	84.67	78.57	91.30
	Specificity	65.03	57.61	41.49	27.62	30.20
	Precision	58.68	35.13	26.64	22.82	28.76
DNN	Sensitivity	73.22	76.31	86.44	66.03	80.00
	Specificity	61.58	52.92	44.60	42.62	47.78
	Precision	56.88	38.53	30.63	25.00	32.18
XGB-1	Sensitivity	81.50	84.37	84.84	55.40	80.00
	Specificity	76.45	51.57	55.35	73.53	66.19
	Precision	64.67	35.43	28.94	30.67	36.03
XGB-2	Sensitivity	81.50	80.00	82.57	53.01	72.00
	Specificity	80.97	58.07	60.06	76.84	74.76
	Precision	69.38	37.53	30.70	32.59	40.44
LSTM	Sensitivity	84.64	88.19	88.63	66.20	84.00
	Specificity	60.79	29.38	35.06	56.74	44.28
	Precision	53.44	28.40	22.63	24.44	26.41

Table 6. Sensitivity for day-ahead, patient-day level code-blue prediction.

ID	03/20-07/20	08/20-09/20	10/20-12/20	12/20-01/21	02/21
LSTM	62.95	65.30	70.58	50.00	78.57
DRF	53.52	30.67	54.36	67.44	64.28
XGB	57.30	60.00	62.10	60.46	78.57
GB	63.27	48.92	67.34	73.52	76.00
DNN	54.77	50.00	62.36	48.48	80.00
XGB-1	46.98	63.26	64.70	61.90	67.85
XGB-2	46.98	57.82	64.70	47.61	57.14

Optimization of Probability Thresholds: For each model, we identified an optimized probability threshold that yielded the highest balanced sensitivity and specificity in the training cohort. Figure 2 plot sensitivity vs specificity for the XGB-2 model, as an example, with the optimal cut-off being the threshold corresponding to the intersection of the sensitivity and specificity curves.

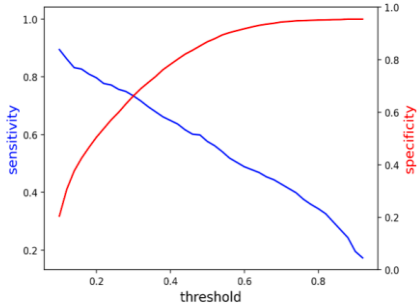

Figure 2. Operating characteristics for the XGB-2 model.

DISCUSSION

Nurse to patient ratios in most healthcare systems range from 5:1 to 10:1, on average a patient's health is typically assessed every 4 to 6 hours, and survival rate after clinical intervention signified by conventional code blue alarms is less than 20% [14]. Centralized continuous monitoring solutions that provide early alerting for ID & Strat of declining health status will enable early intervention by staff [12,13]. Smart EMR systems, precursors to AI, have been shown to reduce patient risk of mortality or morbidity by half and in some cases such systems have accurately identified high-risk events that doctors missed. Some studies report that remote monitoring alone across units dropped ICU mortality by 26% and overall hospital mortality by 16% [11]. In a survey, ICU staff demanded improved alarm management for future patient monitoring systems and clinical decision support systems based on AI was considered useful [15].

Barriers to timely intervention today: To enable proactive intervention at the right moment to avert a catastrophic decline in patient health, clinicians must recognize when a patient is in a high-risk state, which may not be immediately apparent from mere visualization of vital signals. Although central monitoring stations to monitor comprehensive patient information remotely can increase workflow productivity, the lack of independent predictive value based on the correlation of vital signatures and actual outcomes limits the

ability to predict code-blue events sufficiently in advance to stage timely interventions that may avert mortality.

Enhancing current treatment paradigms: Central remote patient monitoring is at best reactionary, even though it reduces time to effective response. By advancing the state of the art to predictive care management, our innovative ID & Strat algorithm has potential to lower mortality rates, reduce readmissions, and improve timeliness of discharge for ICU patients by improving treatment effectiveness and reducing median length-of stay. We anticipate our technology trigger fewer false alarms in the ICU vis-à-vis threshold-driven alerts that are based on individual vital signatures which are less holistic than our multivariate approach, thus reducing burn-out and fatigue amongst first responders and healthcare staff who often find themselves responding to false alarms[15,16]. AI-driven alerts may facilitate early warning systems with the potential to avert adverse events before they occur [17].Other studies have also explored methods to predict code-blue events as well as other high-risk events that prompt urgent clinical attention. In one example, authors have described machine learning models that preemptively flag patients who are likely to go into cardiac arrest and allege performance superior to the *Modified Early Warning Score* used by hospitals [18-20]. The solution we present in this manuscript however is agnostic to specific clinical conditions.

CONCLUSION

We describe a novel machine learning solution that predicts code-blue and other high risk events which require clinical intervention in the ICU starting with patent-specific vital signatures. This gives ICU staff the ability to intervene before the patient becomes critical, thus saving more lives. In our study based on large scale medical time series datasets of patient-specific vital signs in ICU settings from TBH, from a diverse population of adults and pediatric admissions, we were able to achieve out-of-sample sensitivity of over 80% in identifying patients experienced an actionable event occurred in ICUs, with up to 70% sensitivity to patient-specific inpatient days associated with such events. The precision of our models was low and there was variation in performance from month-to-month, however, we anticipate this to improve with the use of a more substantive training dataset. The universe of sensors in ICU beds in this study did not include continuous body temperature monitors. Therefore, the per patient time series vital data acquired for training, validation, and testing lacked a feature that could be informative of a patient's current condition and thus, improve the models' predictive performance. Our long term vision is for care providers, care managers, and payers to utilize the predicted and derived states for a number of health and wellness related applications. We believe that our patient-specific risk scoring models may perform equivalently in other hospital settings than ICU's (eg. ERs, inpatient rooms) and post-discharge wearables-based monitoring scenarios. The latter remains to be explored pursuant to further data collection.

REFERENCES

[1] Haraden, C. (2012). Diagnostics for Patient Safety and Quality of Care. Patient Safety Executive Development Forum. Institute for Healthcare Improvement.

[2] Reyes, J. C. L., Alonso, J. V., Fonseca, J., Santos, M. L., & de los Ángeles Ruiz-Cabello, M. (2016). Characteristics and mortality of elderly patients admitted to the Intensive Care Unit of a district hospital. Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine, 20(7), 391.

[3] Salzman, S. (2019, July 27). How hospitals are using AI to save their sickest patients and curb 'alarm fatigue'. Retrieved Nov 25, 2020.

[4] Lilly, C. M., McLaughlin, J. M., Zhao, H., Baker, S. P., Cody, S., Irwin, R. S., & Group, U. M. C. C. O. (2014). A multicenter study of ICU telemedicine reengineering of adult critical care. Chest, 145(3), 500-507.

[5] Hale, Conor. "FDA Greenlights ICU AI for Predicting Fatal COVID-19 Complications." FierceBiotech, 28 May 2020, www.fiercebiotech.com/medtech

[6] Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, a freely accessible critical care database. Scientific data. 2016;3.

[7] Coustasse, A., Deslich, S., Bailey, D., Hairston, A., & Paul, D. (2014). A business case for tele-intensive care units. The Permanente Journal, 18(4), 76.

[8] Rathore, M. M., Ahmad, A., Paul, A., Wan, J., & Zhang, D. (2016). Real-time medical emergency response system: exploiting IoT and big data for public health. Journal of medical systems, 40(12), 283.

[9] Hayes-Roth, B., Uckun, S., Larsson, J. E., Drakopoulos, J., Gaba, D., Barr, J., & Chien, J. (1994). Guardian: An experimental system for intelligent ICU monitoring. In Proceedings of the Annual Symposium on Computer Application in Medical Care (p. 1004). American Medical Informatics Association.

[10] Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, a freely accessible critical care database. Scientific data. 2016;3.

[11] Mba, Douglas K Richardson, , et al. "A Critical Review of Cost Reduction in Neonatal Intensive Care I. The Structure of Costs." Journal of Perinatology.

[12] Pursley DM, Zupancic JAF. Using Neonatal Intensive Care Units More Wisely for At-Risk Newborns and Their Families. JAMA Netw Open. 2020;3(6):e205693. doi:10.1001/jamanetworkopen.2020.5693.

[13] MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). *DOI:* 10.1038/sdata.2016.35. Available from: http://www.nature.com/articles/sdata201635

[14] Churpek, M. M., Yuen, T. C., Huber, M. T., Park, S. Y., Hall, J. B., & Edelson, D. P. (2012). Predicting cardiac arrest on the wards: a nested case-control study. *Chest*, *141*(5), 1170-1176.

[15] Poncette, A. S., Mosch, L., Spies, C., Schmieding, M., Schiefenhövel, F., Krampe, H., & Balzer, F. (2020). Improvements in patient monitoring in the intensive care unit: survey study. *Journal of Medical Internet Research*, *22*(6), e19091.

[16] Charbonnier S, Gentil S. A trend-based alarm system to improve patient monitoring in intensive care units. Control Eng Pract 2007 Sep;15(9):1039-1050.

[17] Meyer, A., Zverinski, D., Pfahringer, B., Kempfert, J., Kuehne, T., Sündermann, S. H., ... & Eickhoff, C. (2018). Machine learning for real-time prediction of complications in critical care: a retrospective study. *The Lancet Respiratory Medicine*, *6*(12), 905-914.

[18] Somanchi, S., Adhikari, S., Lin, A., Eneva, E., & Ghani, R. (2015, August). Early prediction of cardiac arrest (code blue) using electronic medical records. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 2119-2126).

[19] Yee, C. R., Narain, N. R., Akmaev, V. R., & Vemulapalli, V. (2019). A data-driven approach to predicting septic shock in the intensive care unit. *Biomedical informatics insights*, *11*, 1178222619885147.

[20] Hyland, S. L., Faltys, M., Hüser, M., Lyu, X., Gumbsch, T., Esteban, C., ... & Merz, T. M. (2020). Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature medicine*, *26*(3), 364-373.