INTENSIVE CARE UNIT (ICU) DATA ANALYTICS USING MACHINE LEARNING TECHNIQUES

Zeina Rayan, Marco Alfonse, Abdel-Badeeh M. Salem

***Abstract**: The huge amount of clinical signals and measurements in Intensive Care Units (ICU) will simply overwhelm the person responsible for healthcare support and may cause treatment delays, or clinical errors. This paper analyzes the recent Machine Learning (ML) and Computational Intelligence (CI) techniques which are applied to ICU equipment data in the area of smart health during the period from 2011 till 2018. The results show that ML and CI techniques are used in many ICU applications such as mortality prediction, ICU readmission prediction, prediction of sepsis, prediction of complications in ICU, and processing and monitoring vital signs in ICU patients.*

***Keywords**: Smart Healthcare, Computational Intelligence, Intensive Care Unit, Medical Informatics, Machine Learning.*

***ACM Classification Keywords**: 1.2 Artificial intelligence, F1.1 Models of computation, H.4.2 Types of systems Decision Support.*

Introduction

Towards smart health [Solanas et al, 2014], hospitals infrastructure is now being rebuild, this infrastructure should make use of Information and Communication Technology (ICT). The ICU is designed to care for patients who are seriously injured, have a critical or life-threatening illness, or have undergone a major surgical procedure, therefore requiring 24-hour care and monitoring. The ICUs have been evolved by the evolution of ICT, where an ICU equipment has many devices and sensitive connections between the patient and these devices which leads to better patient monitoring. The ICU equipment includes many devices responsible for various tasks such as patient monitoring, respiratory and cardiac support, emergency resuscitative, pain management, and other life support equipments for people with serious injury. The ICU equipment provides data such as electrical activity of heart, respiratory rate (breathing), blood pressure, body temperature, amount of oxygen and carbon dioxide in the blood, intracranial pressure (pressure of fluid in the brain) … etc. [Johnson et al, 2016].

The continuous monitoring of ICU patients has resulted in enormous amount of data, which leads to many opportunities and challenges [Johnson et al, 2016]. This huge amount of data can be utilized to develop optimal machine learning models that can be used in treatment, diagnosis, and discharging of ICU patients. However, the storage, analysis, privacy, integration, and harmony of these data are considered challenges for handling the data (data analysis and processing).

This paper is constructed as follows: section 2 presents the different machine learning techniques which were applied to ICU data, section 3 presents the comparative study of these techniques and finally section 4 includes the conclusions and future work.

Analysis of recent research ML and CI used in ICU data analytics

This section presents a comparative study between many different machine learning and computational intelligence techniques that are applied to ICU data.

Che, et. al. [Che et al, 2018] proposed a model (GRU_D) which is based on Gated Recurrent Unit (GRU) for mortality prediction using the Multi-Parameter Intelligent Monitoring for Intensive Care, version 3 (MIMIC-III) which is a public research archive collected from different ICU patents [Saeed et al, 2011]. The Proposed model achieved Area Under Curve (AUC) of 0.8527±0.003. In addition, they used the PhysioNet Dataset to validate the proposed model and achieved AUC of 0.8424±0.012.

Nemati, et. al. [Nemati et al, 2018] proposed a model for the prediction of sepsis based on a modified Weibull-Cox proportional hazards model and their dataset was collected from two hospitals within the Emory healthcare system, and other publicly available ICU database. The proposed model achieved 70% accuracy using 4-Hours data of monitoring, 68% accuracy using 6-Hours data of monitoring, 67% accuracy using 8-Hours data of monitoring and 65% accuracy using 12-Hours data of monitoring.

Meyer, et. al. [Meyer el al, 2018] proposed a deep learning model (recurrent deep neural network) for real time prediction of complications in ICU, they used dataset collected from German tertiary care center (German Heart Center Berlin) for cardiovascular diseases. The proposed model achieved accuracy of 86%, then this work was validated using MIMIC-III dataset where the proposed model achieved accuracy of 75%.

Anand, et. al. [Anand et al, 2018] proposed a model for prediction of mortality in diabetic ICU Patients, they used the diabetic inpatients of the MIMIC-III dataset. The proposed model achieved accuracy of 94%.

Viegas, et. al. [Viegas et al, 2017] proposed an ensemble model to predict readmissions of ICU utilizing feature selection and fuzzy modeling approaches. They used MIMIC-II database. The proposed model gives AUC of 0.79.

Desautels, et. al. [Desautels et al, 2016] proposed a model (insight classifier) for prediction of Sepsis in ICU utilizing the MIMIC-III as a dataset. The proposed model achieved accuracy of 80%.

Pirracchio, et. al. [Pirracchio et al, 2015] proposed a model, Super ICU Learner Algorithm (SICULA), for mortality prediction. They trained the model using the MIMIC-III dataset, and The cross-validated area under the receiver operating characteristic curve (AUROC) for the proposed model was 0.85

Leite, et. al. [Leite et al, 2011] proposed a fuzzy model for processing and monitoring vital signs of ICU patients, and used MIMIC as a dataset, where the proposed model achieved accuracy of 96%.

Ramon, et. al. [Ramon et al, 2007] proposed a model for patient survival and length of stay prediction. The proposed model achieved 88% accuracy. They also proposed a model for prediction of development of endangering states where the achieved accuracy is 95%. In addition, they proposed a model for prediction of recovery from endangering states where the achieved accuracy is 87%. The dataset used for these models was collected from various available data sources.

Results and Discussion

This section discusses the different machine learning techniques that are used for the intensive care unit data. These techniques are applied to many tasks such as ICU mortality prediction, ICU sepsis prediction, and ICU monitoring.

Table 1 shows a comparison between different machine learning approaches used in mortality prediction of ICU patients.

Table 1. ML approaches in ICU Mortality Prediction

Authors	Dataset	Preprocessing & Feature Extraction	Machine Learning Technique	Evaluation
Che, et. al., 2018 [Che et al, 2018]	MIMIC-III		Linear Regression	AUC = 0.7715±0.015
		Principle Component Analysis (PCA)	Linear Regression	AUC = 0.7246±0.014
			Support Vector Machine (SVM)	AUC = 0.8146±0.008
		PCA	SVM	AUC = 0.7235±0.012

Authors	Dataset	Preprocessing & Feature Extraction	Machine Learning Technique	Evaluation
			Random Forest (RF)	AUC = 0.8294±0.007
		PCA	RF	AUC = 0.7747±0.009
			GRU	AUC = 0.8380±0.008
			Proposed GRU_D	AUC = 0.8527±0.003
	PhysioNet Dataset		Linear Regression	AUC = 0.7625±0.004
		PCA	Linear Regression	AUC = 0.6890±0.019
			SVM	AUC = 0.8277±0.012
		PCA	SVM	AUC = 0.7741±0.014
			RF	AUC = 0.8157±0.014
		PCA	RF	AUC = 0.7561±0.025

Authors	Dataset	Preprocessing & Feature Extraction	Machine Learning Technique	Evaluation
			GRU	AUC = 0.8226±0.010
			Proposed GRU_D	AUC = 0.8424±0.012
Anand, et. al., 2017 [Anand et al, 2018]	MIMIC-III (The Diabetes inpatients)		RF with threshold 0.04	Accuracy = 61%
			RF with threshold 0.06	Accuracy = 74%
			RF with threshold 0.1	Accuracy = 85%
			RF with threshold 0.12	Accuracy = 89%
			RF with threshold 0.14	Accuracy = 92%
			RF with threshold 0.16	Accuracy = 94%

Authors	Dataset	Preprocessing & Feature Extraction	Machine Learning Technique	Evaluation
Pirracchio, et. al., 2015 [Pirracchio et al, 2015]	MIMIC-II		SAPS II	AUROC = 0.78
			SOFA	AUROC = 0.71
			SICULA	AUROC = 0.85
Ramon, et. al., 2007 [Ramon et al, 2007]	Different available data sources		Decision trees	Accuracy = 79%
			First Order RF	Accuracy = 82%
			Naïve Bayes (NB)	Accuracy = 88%
			Tree Augmented NB	Accuracy = 86%

From table 1, there are many machine learning techniques that can be used for mortality prediction of an ICU patient, such as linear regression, support vector machine, naïve Bayes, …etc. The Random Forest algorithm is the most common one used for mortality prediction.

Table 2 shows a comparison between different machine learning approaches used in prediction of sepsis for patients in ICU.

Table 2. ML approaches in ICU sepsis prediction

Authors	Dataset	Preprocessing & Feature Extraction	Machine Learning Technique	Accuracy
Nemati, et. al., 2018 [Nemati et al, 2018]	Data was collected from two hospitals within the Emory Healthcare system, as well as an external publicly available ICU database	4 hour data	An algorithm is based on a modified Weibull-Cox proportional hazards model	70%
		6 hour data		68%
		8 hour data		67%
		12 hour data		65%
Desautels, et. al., 2016 [Desautels et al, 2016]	MIMIC-III	The most recent bin value is carried forward to fill subsequent empty bins	Insight Classifier	80%
			SIRS	47%
			Quick SOFA	80%
			MEWS	76%
			SAPS II	55%
			SOFA	52%

From table 2, different ML techniques are used for proposing models for sepsis prediction of ICU patients. These models are applied to many patients’ data with different intervals of time. The accuracy achieved for sepsis prediction varies from 47% to 80%.

Table 3 shows a comparison between different machine learning approaches used in ICU patient monitoring for different tasks.

Table 3. ML approaches in ICU monotoring

Authors	**Task**	**Dataset**	**Preprocessing & Feature Extraction**	**Machine Learning Technique**	**Accuracy / AUC**
Meyer, et. al., 2018 [Meyer el al, 2018]	Real time prediction of complications in ICU	Dataset collected from German tertiary care center for cardiovascular diseases from German Heart Center Berlin	The Postoperative bleeding, postoperative renal failure requiring renal replacement therapy, and postoperative in hospital mortality were labeled as “complication occurred” or “complication did not occur”	Recurrent deep neural network	Accuracy = 86%
		MIMIC-III			Accuracy = 75%
Viegas, et. al., 2017 [Viegas et al, 2017]	Prediction of ICU readmissions	MIMIC-II	exactly define the scope of allowable qualities as indicated by the expert	Average ensemble decision criteria	AUC = 0.77±0.02
				Linear SVM	Accuracy = 50.6%
				Polynomial SVM	Accuracy = 52.7%
				Radial basis function based SVM	Accuracy = 69.5%

Authors	Task	Dataset	Preprocessing & Feature Extraction	Machine Learning Technique	Accuracy / AUC
Leite, et. al., 2011 [Leite et al, 2011]	Processing and monitoring vital signs in ICU patients	MIMIC	Extraction of the major physiologic signals that interfere directly in the clinical condition of patients with a stroke diagnosis	Fuzzy Model	Accuracy = 96%
Ramon, et. al., 2007 [Ramon et al, 2007]	Prediction of development of endangering states	Different available data sources		Decision trees	Accuracy = 89%
				First Order RF	Accuracy = 95%
				NB	Accuracy = 95%
				Tree Augmented NB	Accuracy = 92%
	Prediction of recovery from endangering states			Decision trees	Accuracy = 82%
				First Order RF	Accuracy = 87%
				NB	Accuracy = 80%
				Tree Augmented NB	Accuracy = 85%

From table 3, The random forest, naïve Bayes, decision trees, SVM, and deep learning are used for many tasks regarding the ICU data such as complications prediction in ICU, ICU readmissions prediction, prediction of endangering states development, prediction of recovery from endangering states.

Conclusion and Future Work

This work presents an overview of ICU applications, by providing a structured analysis and a comparative study of numerous machine learning techniques which are used for ICU data analysis. The machine learning techniques proven to be a pioneer method for ICU data analysis. The MIMIC dataset is the most commonly used as a source of ICU data. In future work, we are going to apply the machine learning algorithms to anomaly detection over ICU data.

Bibliography

[Anand et al, 2018] Anand, R.S., Stey, P., Jain, S., Biron, D.R., Bhatt, H., Monteiro, K., Feller, E., Ranney, M.L., Sarkar, I.N. and Chen, E.S., Predicting Mortality in Diabetic ICU Patients Using Machine Learning and Severity Indices. AMIA Summits on Translational Science Proceedings, p.310, 2018.

[Che et al, 2018] Che, Z., Purushotham, S., Cho, K., Sontag, D. and Liu, Y. Recurrent neural networks for multivariate time series with missing values. Nature scientific reports, 8(1), p.6085, 2018.

[Desautels et al, 2016] Desautels, T., Calvert, J., Hoffman, J., Jay, M., Kerem, Y., Shieh, L., Shimabukuro, D., Chettipally, U., Feldman, M.D., Barton, C. and Wales, D.J. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR medical informatics, 4(3), 2016.

[Johnson et al, 2016] Johnson, A.E., Ghassemi, M.M., Nemati, S., Niehaus, K.E., Clifton, D.A. and Clifford, G.D. Machine learning and decision support in critical care. Proceedings of the IEEE. Institute of Electrical and Electronics Engineers, 104(2), pp.444-466, 2016.

[Leite et al, 2011] Leite, C.R., Sizilio, G.R., Neto, A.D., Valentim, R.A. and Guerreiro, A.M. A fuzzy model for processing and monitoring vital signs in ICU patients. Biomedical engineering online, 10(1), p.68, 2011.

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

[Nemati et al, 2018] Nemati, S., Holder, A., Razmi, F., Stanley, M.D., Clifford, G.D. and Buchman, T.G. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Critical care medicine, 46(4), pp.547-553, 2018.

[Pirracchio et al, 2015] Pirracchio, R., Petersen, M.L., Carone, M., Rigon, M.R., Chevret, S. and van der Laan, M.J. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. The Lancet Respiratory Medicine, 3(1), pp.42-52, 2015.

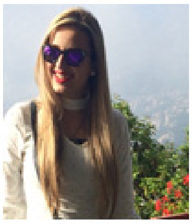
[Ramon et al, 2007] Ramon, J., Fierens, D., Güiza, F., Meyfroidt, G., Blockeel, H., Bruynooghe, M. and Van Den Berghe, G. Mining data from intensive care patients. Advanced Engineering Informatics, 21(3), pp.243-256, 2007.

[Saeed et al, 2011] M. Saeed, M. Villarroel, A.T. Reisner, G. Clifford, L. Lehman, G.B. Moody, T. Heldt, T.H. Kyaw, B.E. Moody, R.G. Mark. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access ICU database. Critical Care Medicine 39(5):952-960; doi: 10.1097/CCM.0b013e31820a92c6, 2011.

[Solanas et al, 2014] Solanas, A., Patsakis, C., Conti, M., Vlachos, I.S., Ramos, V., Falcone, F., Postolache, O., Pérez-Martínez, P.A., Di Pietro, R., Perrea, D.N. and Martinez-Balleste, A. Smart health: a context-aware health paradigm within smart cities. IEEE Communications Magazine, 52(8), pp.74-81, 2014.

[Viegas et al, 2017] Viegas, R., Salgado, C.M., Curto, S., Carvalho, J.P., Vieira, S.M. and Finkelstein, S.N., Daily prediction of ICU readmissions using feature engineering and ensemble fuzzy modeling. Expert Systems with Applications, 79, pp.244-253, 2017.

Authors' Information



Zeina Rayan – *T.A. & Msc Student, Computer Science Department, Faculty of Computer and Information Scinces, Ain Shams University, Cairo, Egypt;*

e-mail: *zeinarayan@ cis.asu.edu.eg*

Major Fields of Scientific Research: *Artificial Intelligence, Machine Learning, Smart Health, Medical Informatics*

Marco Alfonse – *Lecturer at faculty of computer and information sciences, Ain Shams University, Abbasia, Cairo, Egypt,*

e-mail: marco@fcis.asu.edu.eg

Major Fields of Scientific Research: *Semantic Web, Ontological Engineering, Medical Informatics, Artificial Intelligence, Expert Systems, Image Processing, Knowledge Engineering*

Abdel-Badeeh M. Salem *– Professor at faculty of computer and information sciences, Ain Shams University, Abbasia, Cairo, Egypt,*

e-mail*: absalem@cis.asu.edu.eg*

Major Fields of Scientific Research*: Artificial intelligence(AI), Computational intelligence, Machine learning, Knowledge engineering, Big data analytics, Intelligent biomedical informatics and Healthcare systems, and AI education*