Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200125

Blood Lactate Concentration Prediction in Critical Care

Behrooz MAMANDIPOOR[1] [a], Mahshid MAJD [a], Monica MOZ [a,b] and Venet OSMANI [a]
[a] *Fondazione Bruno Kessler Research Institute, Trento, Italy*
[b] *Humanitas Research Hospital, Milan, Italy*

Abstract. Blood lactate concentration is a reliable risk indicator of deterioration in critical care requiring frequent blood sampling. However, lactate measurement is an invasive procedure that can increase risk of infections. Yet there is no clinical consensus on the frequency of measurements. In response we investigate whether machine learning algorithms can be used to predict blood lactate concentration from ICU health records. We evaluate the performance of different prediction algorithms using a multi-centre critical care dataset containing 13,464 patients. Furthermore, we analyse impact of missing value handling methods in prediction performance for each algorithm. Our experimental analysis show promising results, establishing a baseline for further investigation into this problem.

Keywords. machine learning, deep learning, clinical decision support, critical care

1. Introduction

Blood lactate concentration is frequently measured in critical care as it is used as an indicator of risk of patients' deterioration. Presence of increased lactate levels is correlated with increased risk of morbidity and mortality. In healthy individuals there is a continuous cycle of lactate production and clearance (metabolised primarily in the liver), while in critically ill patients lactate metabolism is impaired, resulting in elevated lactate concentration. Several clinical conditions have been associated with impaired clearance of lactate, such as liver dysfunction and sepsis. Therefore frequent lactate measurements are necessary to track and assess patients' state. However, lactate concentration cannot be measured without drawing arterial or venous blood, which is an invasive procedure that can increase risk of infections. As a consequence, blood gas analysis may not be ordered as frequently leading to sub optimal rates of lactate measurements [1].

In this respect, there is ample potential for machine learning methods to play a significant role in lactate guided clinical decision making. Such role is especially important when considering that lactate-guided therapy significantly reduced hospital mortality and length of stay as evidenced by several multicentre randomised controlled trials [2,3,4]. However, this potential has not been explored thus far. Addressing the evident need for prediction of blood lactate concentration in ICU, we are the first to compare performance

[1] Corresponding Author: E-mail: bmamandipoor@fbk.eu

of several machine learning methods in lactate concentration prediction. Furthermore, we evaluate several strategies to handle missing values using a dataset of 13,464 patients, containing 12,196,798 clinical records from the eICU critical care database [5]. Our results show that blood lactate concentration prediction could become a useful tool during clinical decision making process in critical care, that may reduce unnecessary blood sampling, while meeting recommendations of Surviving Sepsis Campaign (SSC) on serial lactate testing [6].

2. Methods

We formally define the problem of lactate concentration prediction as follows: For each patient with the set of clinical parameters $S = \{(l, X)_t; t \in T\}$, where l represents blood lactate concentration, X represents the set of all other clinical measures, and t is the time index, we want to predict $l_{t+\beta}$ based on $\hat{S}_t = \{l, \widehat{X})_{t'} \in [t-\alpha : t]\}$, where $\hat{X}$ is the selected set of clinically relevant measurements out of X. Therefore, we formulate a regression problem using Eq.(1).

$$f: \hat{S}_t \rightarrow l_{t+\beta}\ , \min_{\theta} Loss\left(f\left(\hat{S}_t, \theta\right), l_{t+\beta}\right) \tag{1}$$

In other words, the objective of lactate prediction is to predict the blood lactate concentration of a patient in the next β hours using a selected set of their measurements taken in the past α hours.

For our work we use eICU Collaborative Research Database [5], a multi-center intensive care unit database with high granularity data. The eICU database comprises 200,859 patient unit encounters for 139,367 unique patients admitted between 2014 and 2015 to hospitals located throughout the US. The final patient cohort contained 13,464 patients (14,477 ICU stays) with 12,196,798 clinical records, where we grouped these records into 2 hour windows. Patient's mean age was 61.8 ± 15.7 years (45% female). For our case study, the relevant clinical features for the selected cohort are chosen based on advice from the clinician and outlined in Table 1.

Critical care data suffers from several limitations with missing data as one of the main challenges. In the eICU dataset, the percentage of missingness per feature is between 0 to 96 which needs to be addressed properly. Addressing missing values is typically dependent on the cause of missingness. Statistically speaking, there are three types of missing values: missing completely at random (MCAR) which happens on an unrelated cause, missing at random (MAR) which implies a relation between missing value and other present values, and missing not at random (MNAR) which implies a relationship between the value of the variable and its missingness [7]. On the other hand, there are several imputation methods to handle missingness. Mean and forward imputations are basic single-value imputation methods, which only consider the information in the past values of the variable and work best under MCAR assumption. Multiple imputation [8], Matrix Factorisation [9], PCA [10], SoftImpute [11] and Random Forest [12] are the most known traditional machine learning methods that find the substitute for missing values based on the relation between observed and missing features. These methods perform best when handling MAR cases. Finally, MNAR cases are the hardest to manage. A helpful solution to capture the information behind missingness of variables is to use missing

Table 1. List of selected variables from eICU tables (in bold) based on clinical relevance

Patient, gender, age, ethnicity, admissionweight, apacheadmissiondx
Lab, Respiratory Rate, O2 Saturation, FiO2, glucose, potassium, sodium, Hgb, chloride, creatinine, BUN, bicarbonate, LPM O2, calcium, Hct, platelets x 1000, anion gap, WBC x 1000, lactate, RBC, RDW, paO2, pH, paCO2, magnesium, HCO3, Total CO2, Base Excess, phosphate, Pressure Support, albumin, -lymphs, -polys, -eos, -basos, -bands, -monos, bilirubin, AST (SGOT), ALT (SGPT), total protein, alkaline phos.
NurseCharting, Heart Rate, Respiratory Rate, Temperature (C), Invasive BP Mean, Invasive BP Systolic, Invasive BP Diastolic, Non-Invasive BP Mean, Non-Invasive BP Systolic, Non-Invasive BP Diastolic, O2 Saturation, glucose, CVP, LPM O2, Total CO2
RespiratoryCharting, Heart Rate, Respiratory Rate, FiO2, Total CO2, Tidal Volume, Inspiratory Pressure, LPM O2, Vent Rate, Plateau Pressure, Mean Airway Pressure, Pressure Support, Inspiratory Pressure
VitalPeriodic, Heart Rate, Respiratory Rate, Temperature (C), Invasive BP Mean, Invasive BP Systolic, Invasive BP Diastolic, CVP
VitalAperiodic, Non-Invasive BP Mean, Non-Invasive BP Systolic, Non-Invasive BP Diastolic
IntakeOutput, Urine

indicators. These indicators can either be used directly aside data or inside the prediction model structure. Note that indicators can also be useful in MAR and MCAR, since they differentiate the imputed values from the observed ones and the prediction method can use this information to ignore bad imputations.

There are a vast number of successful regression methods used for healthcare data which could be grouped into three categories: the statistical regression methods, the traditional machine learning methods and the rising deep learning methods. Here we investigate one method from each category: Lasso regression(LR) as linear regression, to benefit from its feature section capabilities, Random Forest(RF) as nonlinear model with an ensemble of decision trees, and Long Short-term Memory(LSTM) as temporal deep learning model to learn complex temporal relation in data.

3. Results

Data preparation. Data is prepossessed in six steps. (1) Patient cohort is selected based on three inclusion criteria; adult patients (age > 18), with at least two measured lactate levels, and with at least 18 hours length of stay in ICU. (2) The relevant variables are selected based on advice from the clinician. For our case study, the relevant clinical features for the selected cohort are outlined in Table 1. A detailed description of each variable is provided in the original eICU paper [5]. (3) Some features exist in more than one table under different names. These features are aligned to a unique feature in time. (4) The selected data is aligned in time since each feature is measured in an arbitrary time and frequency. We resampled time-series data into regularly aligned periods where each feature is sampled every 2 hours. In case a feature is measured more than once during each two hour interval, the last record is used. (5) Noise and outliers are addressed as follows: for each feature, the valid interval is defined based on clinical knowledge and the values out of the valid scope are considered as missing values. (6) The data is split into training and testing parts, which is done using five-fold cross validation where at each fold 80% of the data is considered as training and the rest is test data.

Settings. All the imputation methods are applied using python based on

fancyimpute, predictive imputer, and sklearn open-source libraries. Also for all the machine learning methods, the default parameters proposed by their authors are used. The LSTM network has 2 layers of 1024 units with *Glorot* normalisation and *tanh* activation, each followed by a drop out layer of 0.6. Adam optimizer is used with learning rate starting from 0.0001 and the model is trained for 20 epochs with batch size 100. To ease LSTM convergence, data is normalised to have zero mean and a standard deviation of one. The LSTM model is implemented using Keras and Tensorflow as backend and one GTX 2080 Ti as GPU.

All combinations of introduced prediction models and imputation methods are examined on eICU data and their results are reported in Table 2. To measure the quality of imputation methods regarding lactate prediction while preserving the structure of data, we report Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) as indicators of the predictive performance of the regression models. The mean and standard deviation of each measure on five fold cross-validation of data is reported. The best results are shown in bold. As the results suggest, both regression model and imputation method affect the prediction results. LSTM performed significantly better compared to LR an RF using the same imputation method. Therefore, it can be concluded that the data contains complex relations not only between measurements but also across their values in time that were captured by the LSTM model.

Table 2. The results of all combinations of prediction models and imputation methods. The mean and standard deviation of various metrics on five fold cross validation of data is provided. MAE, RMSE, and R^2 stand for Mean Absolute Error, Root Mean Squared Error, and R-squared respectively. LR, RF, and LSTM stand for Linear Regression, Random Forest, and Long Short Term Memory.

Measure	Imputation \ Regression	LR	RF	LSTM
MAE	Mean	0.859 ± 0.006	0.856 ± 0.007	0.745 ± 0.045
	Feed Forward	0.735 ± 0.008	0.733 ± 0.009	0.692 ± 0.008
	Indicator	0.725 ± 0.009	0.720 ± 0.010	**0.665 ± 0.009**
	PCA	0.864 ± 0.006	0.861 ± 0.008	0.712 ± 0.008
	MF	0.905 ± 0.009	0.901 ± 0.010	0.715 ± 0.012
	SoftImpute	0.862 ± 0.010	0.858 ± 0.011	0.705 ± 0.006
	MissForest	0.853 ± 0.008	0.849 ± 0.005	0.714 ± 0.010
	MICE	0.872 ± 0.007	0.869 ± 0.008	0.715 ± 0.009
	AE	0.846 ± 0.007	0.845 ± 0.008	0.730 ± 0.051
RMSE	Mean	1.263 ± 0.013	1.257 ± 0.015	1.120 ± 0.016
	Feed Forward	1.120 ± 0.014	1.115 ± 0.015	1.075 ± 0.016
	Indicator	1.090 ± 0.013	1.085 ± 0.016	**1.016 ± 0.025**
	PCA	1.268 ± 0.014	1.262 ± 0.016	1.104 ± 0.014
	MF	1.304 ± 0.015	1.298 ± 0.016	1.100 ± 0.018
	SoftImpute	1.268 ± 0.016	1.264 ± 0.017	1.095 ± 0.017
	MissForest	1.254 ± 0.016	1.248 ± 0.011	1.100 ± 0.011
	MICE	1.270 ± 0.014	1.266 ± 0.015	1.105 ± 0.016
	AE	1.245 ± 0.013	1.241 ± 0.015	1.103 ± 0.022
R^2	Mean	0.475 ± 0.010	0.480 ± 0.013	0.585 ± 0.022
	Feed Forward	0.587 ± 0.008	0.591 ± 0.011	0.620 ± 0.013
	Indicator	0.605 ± 0.012	0.610 ± 0.013	**0.660 ± 0.017**
	PCA	0.471 ± 0.011	0.476 ± 0.014	0.598 ± 0.011
	MF	0.440 ± 0.013	0.445 ± 0.015	0.601 ± 0.014
	SoftImpute	0.471 ± 0.010	0.475 ± 0.013	0.605 ± 0.006
	MissForest	0.482 ± 0.007	0.487 ± 0.006	0.601 ± 0.012
	MICE	0.469 ± 0.010	0.472 ± 0.012	0.598 ± 0.011
	AE	0.489 ± 0.010	0.493 ± 0.012	0.599 ± 0.024

4. Conclusion

Considering the importance of monitoring blood lactate levels, the ability to predict blood lactate concentration may provide clinicians sufficient time to devise interventions for rapidly deteriorating patients. At the same time it may decrease the frequency (and the associated cost) of invasive lactate measurements in stable patients. Therefore, we defined the problem of lactate prediction from ICU data in details and offered solutions with well-known prediction and imputation methods.

Our results show that LSTM-based method can predict lactate level with a Mean Absolute Error of 0.665 across 13,464 patients from different hospitals and ICU units. Furthermore, we show that indicator imputation method achieves highest performance in our dataset, suggesting that a missing value indicator is informational and increases predictive power over other, mean-based imputation methods. Future work will investigate prediction of lactate levels in patient sub-populations and its impact on critical care at patient as well as organisational level.

This work was partially supported by the European Union's Horizon 2020 research and innovation program under grant agreement #769765.

References

[1] C. Rhee, M. V. Murphy, L. Li, R. Platt, M. Klompas, Lactate testing in suspected sepsis: trends and predictors of failure to measure levels, Critical care medicine 43 (8) (2015) 1669 (2015).

[2] T. C. Jansen, J. van Bommel, F. J. Schoonderbeek, S. J. Sleeswijk Visser, J. M. van der Klooster, A. P. Lima, S. P. Willemsen, J. Bakker, Early lactate-guided therapy in intensive care unit patients: a multicenter, open-label, randomized controlled trial, American journal of respiratory and critical care medicine 182 (6) (2010) 752–761 (2010).

[3] A. E. Jones, N. I. Shapiro, S. Trzeciak, R. C. Arnold, H. A. Claremont, J. A. Kline, E. M. S. R. N. E. Investigators, et al., Lactate clearance vs central venous oxygen saturation as goals of early sepsis therapy: a randomized clinical trial, Jama 303 (8) (2010) 739–746 (2010).

[4] H. Tian, S. Han, C. Lv, T. Wang, Z. Li, D. Hao, Q. Shang, X. Wang, The effect of early goal lactate clearance rate on the outcome of septic shock patients with severe pneumonia, Zhongguo wei zhong bing ji jiu yi xue= Chinese critical care medicine= Zhongguo weizhongbing jijiuyixue 24 (1) (2012) 42–45 (2012).

[5] T. J. Pollard, A. E. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, O. Badawi, The eicu collaborative research database, a freely available multi-center database for critical care research, Scientific data 5 (2018).

[6] M. M. Levy, L. E. Evans, A. Rhodes, The surviving sepsis campaign bundle: 2018 update, Intensive care medicine 44 (6) (2018) 925–928 (2018).

[7] D. B. Rubin, Inference and missing data, Biometrika 63 (3) (1976) 581–592 (1976).

[8] M. J. Azur, E. A. Stuart, C. Frangakis, P. J. Leaf, Multiple imputation by chained equations: what is it and how does it work?, International journal of methods in psychiatric research 20 (1) (2011) 40–49 (2011).

[9] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer (8) (2009) 30–37 (2009).

[10] J. Josse, F. Husson, Handling missing values in exploratory multivariate data analysis methods, Journal de la Société Française de Statistique 153 (2) (2012) 79–99 (2012).

[11] R. Mazumder, T. Hastie, R. Tibshirani, Spectral regularization algorithms for learning large incomplete matrices, Journal of machine learning research 11 (Aug) (2010) 2287–2322 (2010).

[12] D. J. Stekhoven, P. Bü hlmann, Missforest—non-parametric missing value imputation for mixed-type data, Bioinformatics 28 (1) (2011) 112–118 (2011).