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

Length of Stay in the Neonatal ICU is Predictable using Heart Rate: An Opportunity for Optimizing Managed Care

Xinyu (Ivy) Zhang, Prahlad G Menon

***Abstract*— We explore the use of classification and regression models for predicting the length of stay (LoS) of neonatal patients in the intensive care unit (ICU), using heart rate (HR) time-series data of 7,758 patients from the MIMIC-III database. We find that aggregated features of HR on the first full-day of in-patient stay after admission (i.e. the first day with a full 24-hour record for each patient) can be leveraged to classify LoS in excess of 10 days with 89% sensitivity and 59% specificity. As such, LoS as a continuous variable was also found to be statistically significantly correlated to aggregate HR data corresponding to the first full-day after admission.**

I. BACKGROUND

The costs of neonatal ICU (NICU) 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 NICU utilization in order to optimize management of care. However, while costs of care are known to have a direct relationship with length of stay (LoS, in days) and has been shown to have an inverse relationship with gestational age, with extreme preterm cases being of the highest risk, there is still a need for predictive models that can be utilized to exercise optimal strategies for staffing key healthcare personnel and to estimate the impact of treatment approaches as well as identify patient care management paradigms that correlate with a reduced LoS [2,3]. Further, heart rate (HR) is a parameter that is universally measurable as a time-series using wearable devices and in hospital as well as telemedicine settings. Therefore, in this study, we sought to leverage HR and derived features of these time-series data in order to predict LoS in the NICU.

II. METHODS

We utilized the Medical Information Mart for Intensive Care (MIMIC-III) dataset [4] as a source of time-series data and associated ground-truth outcomes data for this study. MIMIC-III is a large, single-center database comprising information on patients admitted to ICUs and includes 7,758 patients (i.e. 83,019 records) in the neonatal age range. We identified heart rate (HR) time-series data of all patients during their stay in the ICU unit, and extracted aggregate features of HR data on a per-visit-per-day basis to compute a set of aggregated features, including mean, median, standard deviation, min, max, skew and the first and last values of HR for each ICU visit, per-day, as well as the slope of HR over time in seconds since the first available record for a given ICU visit. All features were normalized between 0 and 1 across all available records.

First, we established a classification model for LoS in excess of the mean LoS (i.e. 10.7 days). LoS>10 days accounted for 27% of the total available visits in the dataset. This corresponded to 85% of the available aggregated records per-visit-per-day. A series of classification models were built using Naïve Bayes, Artificial Neural Networks, Generalized Linear Modeling (i.e. Logistic Regression) and Random Forest classification, training to predict the binary response of whether the neonatal ICU visit lasted in excess of 10 days (i.e. LoS > 10). The top model from a competition of models was selected based on the basis of sensitivity in detecting long inpatient stays, based on a standard 50% probability threshold for dichotomizing model-specific probability estimates.

In order to develop train and testing sets for modeling while ensuring continuity of visit-specific records, we first split the data corresponding to unique NICU visits into two sets, in a 70% training and 30% testing split. Two classification models were trained, the first one to predict the response variable using the features of HR on the all days after admission, and a second one to predict the response variable using features of HR on the first full-day after admission only.

TABLE I. NUMBER OF NICU VISITS IN TRAINING

Model #	*Data utilized*	*Numbers of Visits*
1	All days	5,430
2	1st full-day only	2,934

TABLE II. NUMBER OF NICU VISITS IN TESTING

Test sets	*Numbers of Visits*
Day of admission	2,328
1st full-day	1,246
2nd full-day	1,024
3rd full-day	921
4th full-day	866

To train the second model, we utilized the records from the training split considering data corresponding to only a specific integer number of days after admission to conduct training and testing respectively i.e. the visit-specific status of LoS > 10 days was trained to be predicted based on the first complete day of inpatient stay after admission. Further, patients considered in the training sets in each run were only those which had at least a complete 24-hours of that day; for example, in order to develop a model based on the first full-day of inpatient stay, we select patients that spent at least one

Xinyu (Ivy) Zhang is a student with the department of statistics at Southwestern University of Finance and Economics (e-mail: ivyzzzhang@gmail.com).

Prahlad G Menon, PhD is a professor with the department of bioengineering at University of Pittsburgh (phone: 412-259-3031; e-mail: prm44@pitt.edu).

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

full-day in the ICU, and utilize only computed features corresponding with that day after admission for training.

To test these two trained models, we developed five independent test sets from the 30% test split, composed of records from the day of admission, the 1st full-day, and so on until the 4th full-day after admission, respectively. The total numbers of selected NICU visits and days used for training and testing as described above are listed in Tables 1 and 2. The number of visits available in our dataset was inversely related to the duration of stay i.e. longer durations of stay were less frequent than short stays in the NICU. In addition to our LoS > 10 days classification model, we developed regression models to estimate LoS as a continuous response variable. For regression modeling, we utilized the same two training sets and five test sets shown in Tables 1 and 2. Further, regression models were fit iteratively for each training set. First we considered all features computed and estimated LoS by visit and by day. Next, we adopted a stepwise backward elimination approach for model fitting where initially all features were adopted and subsequently variables were eliminated in each iteration, in order to optimize R^2 while minimizing variance inflation (i.e. VIF < 10) and selecting regressors only with statistically significant coefficient estimates (alpha = 0.05).

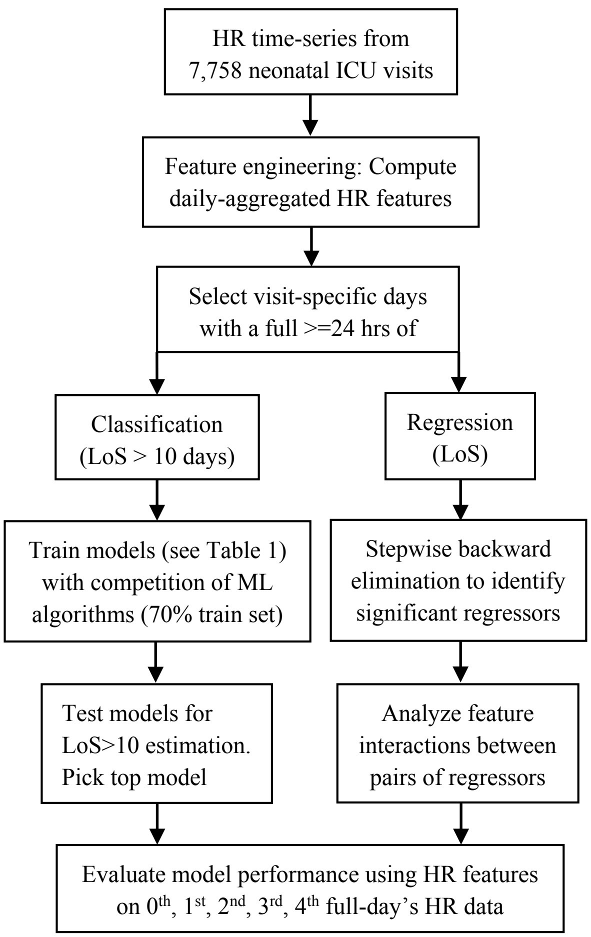

Figure 1. Flowchart of methods adopted for classification and regression models developed in this study. For each i.e. the classification and regression studies we had two sets of models: a) training using only the 1st full-day of admission; and b) training with all days of length of stay.

Finally, we adopted the key features identified from the stepwise backward process along with their pair-wise interaction between these variables in order to create our final model. In testing, R^2 and mean squared error (MSE) from each test set were utilized to gauge performance. Figure 1 summarizes our approach.

III. Results

Random Forest classification led to the best classification performance in terms of sensitivity in identifying long LoS (i.e. LoS > 10 days) in the NICU, with the best model-specific performance emerging based on the aggregated HR data from the 1st full-day of admission. We observed a downward trend in sensitivity after the 1st full-day of time-series data (starting each day at 00:00 hrs) but specificity was observed to increase using data from days further into the duration of stay of a given patient. These trends were found to be consistent even beyond the 5th day of admission (see Figures 2).

HR on the 1st full-day of admission was found to optimal for classifying LoS>10 days, with 89% sensitivity and 59% specificity when training with all-day records, using a 50% probability threshold on predictions. There is scope to optimize this threshold using receiver operator characteristics analysis in order to achieve a balanced sensitivity and specificity of ~74%.

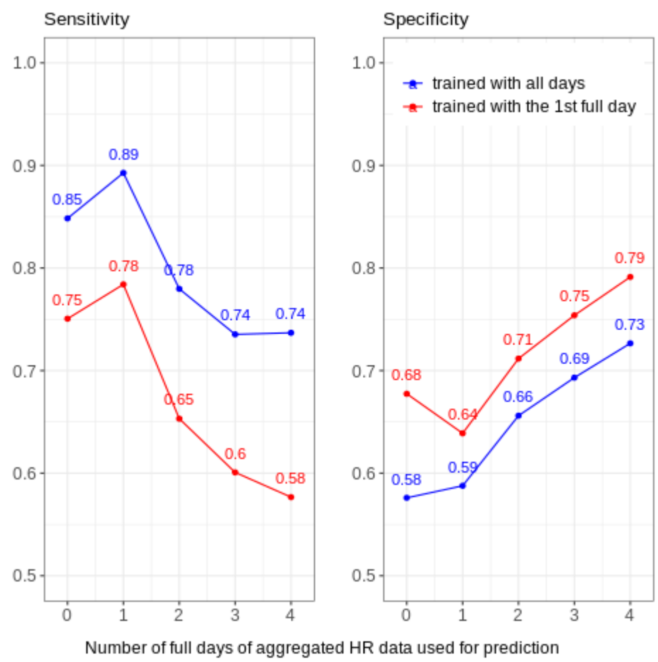

Figure 2. Sensitivity (LEFT) and specificity (RIGHT) in out-of-sample test sets of the two classification modeling approaches adopted in this study to predict LoS > 10 days, starting with aggregated HR data from the 0th, 1st, 2nd, 3rd and 4th full-days after patient admission.

The precision for estimating a long LoS decreased (i.e. the false positive rate increased) using HR data further into the duration of admission. However, the AUC was found to improve (i.e. overall classification accuracy improved) further into the duration of admission in the model that utilized data from all days of admission for training. AUC was conversely found to diminish when using HR further into the duration of stay in the model trained only using HR on the 1st full-day of admission, justifiably owing to its training bias.

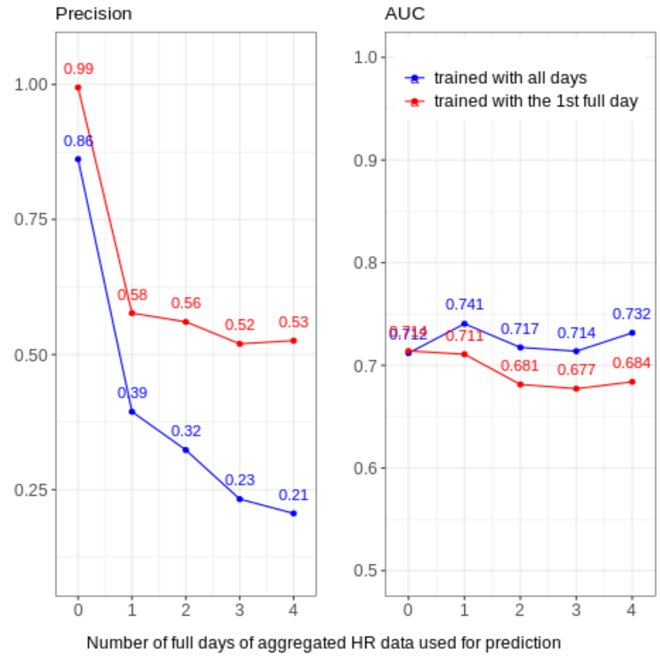

Figure 3. Precision (LEFT) and area under curve (AUC, RIGHT) in out-of-sample test sets of the two classification modeling approaches adopted in this study to predict LoS > 10 days, starting with aggregated HR data from the 0th, 1st, 2nd, 3rd and 4th full-days after patient admission.

Results of the regression modeling of LoS pursuant to features selection once all variables causing variance inflation were eliminated, resulted in an R^2 of 0.21 ($p < 0.01$), when trained with all days of aggregated HR data. The features selected into these models were mean, standard deviation, max, min, high, time since the first record and the average time elapsed between consecutive records. Accounting for selected feature interactions led to an increase in R^2 to 0.24 ($p < 0.01$).

TABLE III. REGRESSION PERFROMANCE OF MODELS TRAINED WITH ALL DAYS. CORRELATIONS WERE STATISTICALLY SIGNIFICANT (P<0.01) .

	R^2	*Adjusted* R^2
Baseline	0.2098	0.2096
Baseline with feature selection	0.209	0.2089
With feature interacions	0.2388	0.2384

Results pursuant to features selection resulted in an R^2 of 0.18 ($p < 0.01$), when trained with the 1st full-days of aggregated HR data. The features selected into these models are mean, standard deviation, the average time elapsed between consecutive records, last record of heart rate, trend of heart rate and time since the first record. Accounting for selected feature interactions led to an increase in R^2 to 0.21 ($p < 0.01$).

Regression using all days of data consistently had better out-of-sample performance relative to using the first full-day of data alone. Using the data from the 0th full-day to the 4th full-day after admission for testing, Figure 3 plots R^2 and means square error (MSE) as indicators of test performance for each day's HR features, independently. MSE was found to increase using HR data day by day further into a given NICU visit, until 5 days into a given admission. Given that MSE was lowest using the 1st full-day's HR data (the 0th full day is ignored), the 1st day after admission may be ideal for estimating LoS accurately.

TABLE IV. REGRESSION PERFROMANCE OF MODELS TRAINED WITH 1ST FULL-DAY. CORRELATIONS WERE STATISTICALLY SIGNIFICANT (P<0.01).

	R^2	*Adjusted* R^2
Baseline	0.1817	0.1783
Baseline with feature selection	0.1806	0.1789
With feature interactions	0.2059	0.2002

Similarly, comparing R^2 results on the test set revealed that HR on the 1st or 2nd full-day after admission was associated with the highest R^2.

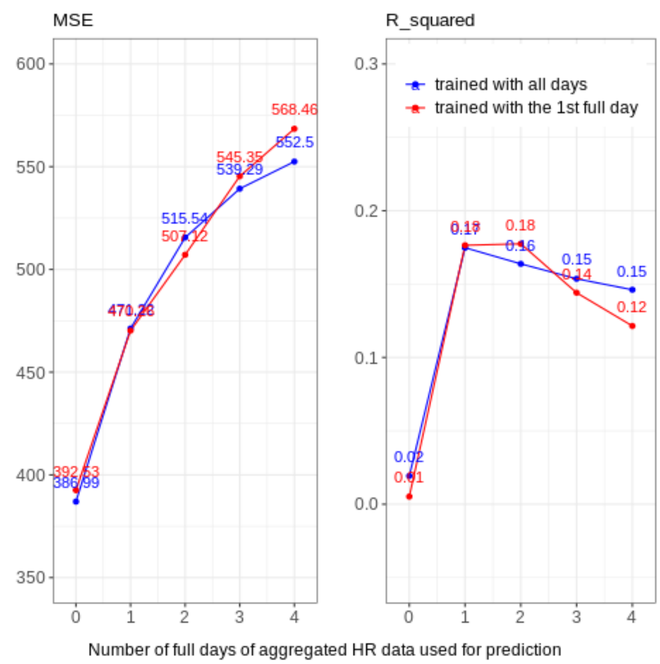

Figure 4. MSE (LEFT) and R^2 (RIGHT) of the two regression modeling approaches adopted in this study to compare the performance in out-of-sample test sets, starting with aggregated HR data from the 0th, 1st, 2nd, 3rd and 4th full-days after patient admission.

Using the data from the 0th full-day to the 99th full-day after admission for testing, Figure 4 plots MSE and number of NICU visits in testing changes with the number of full days used for prediction. MSE was found to increase in the first few days and gradually decrease while number of NICU visits keep decreasing. The figure shows that the aggregated HR data in the first 5 days after admission have better predictive power in determining LoS, and the first full-day is the most predictive. However, in the cases with extremely long LoS (eg: LoS > 60 days), MSE in LoS estimation was found to diminish further than even that achieved with HR on the first day of admission, presumably owing to the fact that a patient's prognosis becomes clearer after a certain duration of stay. However, the latter requires further exploration in a larger cohort of patients, given that MIMIC-III had fewer than ~250 patients per unique count of days of admission that were admitted for a length of stay in excess of 30 days.

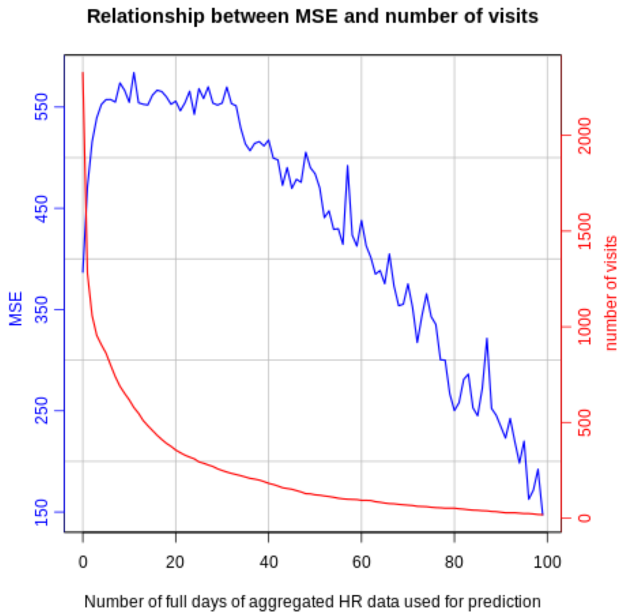

Figure 5. MSE (BLUE) and Number of visits (RED) of the regression modeling approaches trained with all days to show the relationship between MSE and number of visits in the long-term performance of out-of-sample test sets, starting with aggregated HR data from the 0th, 1st, 2nd, 3rd until 99th full-days after patient admission.

IV. Discussion & Conclusion

Our study based on NICU patient data in the MIMIC-III database determines that LoS is statistically significantly predictable using HR on the 1st full-day of admission and further that long-duration LoS is possible to be determined accurately as early as after the 1st full-day of NICU stay after admission. However, as the patient remains in the ICU, the HR no longer independently as predictive of the LoS anymore.

In our out-of-sample testing exercises, we evaluated each full-day of HR data available since the day of admission for each NICU visit for independent predictive value in determining long LoS (i.e. LoS > 10 days, as defined in this study). However, in future, it may be feasible to create an aggregated index of long LoS, which aggregates the predictions of our model from different days of admission, in order to serve healthcare administrators with a single visit-specific index of an expectation for a long LoS.

Our findings are congruent with other studies on NICU LoS [5] but additionally offer an understanding of specifically the predictive power of specifically HR features aggregated on a daily basis in estimating LoS for care management purposes as early as the first full-day after admission.

While it is arguable that predicting premature births [6] is more clinically relevant than predicting and reducing LoS from the perspective of managing care of neonates, we believe that better estimation of LoS using predictive models underpinned on real-time, patient-specific vital / physiological signals, will help with management of staff and resources in the NICU more efficiently from the perspective of the payer and optimization of utilization management.

References

[1] Cheah, Irene Guat Sim. "Economic assessment of neonatal intensive care." Translational pediatrics vol. 8,3 (2019): 246-256. doi:10.21037/tp.2019.07.03

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

[3] Marbella AM, Chetty VK, Layde PM. Neonatal hospital lengths of stay, readmissions, and charges. Pediatrics. 1998 Jan;101(1 Pt 1):32-6. doi: 10.1542/peds.101.1.32. PMID: 9417147.

[4] 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.

[5] Bender, G., Koestler, D., Ombao, H. et al. Neonatal intensive care unit: predictive models for length of stay. J Perinatol 33, 147–153 (2013). https://doi.org/10.1038/jp.2012.62.

[6] 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.