Admission Duration Model for Infant Treatment (ADMIT)

Keith Feldman
Dept. of Computer Science & Eng., iCeNSA
University of Notre Dame, USA
Email: kfeldman@nd.edu

Nitesh V. Chawla
Dept. of Computer Science & Eng., iCeNSA
University of Notre Dame, USA
Email: nchawla@nd.edu

***Abstract*—In today's healthcare environment, nurses play an integral role in determining patient outcomes. This role becomes especially clear in intensive care units such as the Neonatal Intensive Care Unit (NICU). In the NICU, critically ill infants rely almost completely on the care of these nurses for survival. Given the importance of their role, and the volatile conditions of the infants, it is imperative that nurses be able to focus on the infants in their charge. In order to provide this level of care there must be an appropriate infant to nurse ratio each day. However traditional staffing models often utilize a number of factors, including historical census counts, which when incorrect leave a NICU at risk of operating barely reaching, or even below the recommended staffing level. This work will present the novel ADMIT (Admission Duration Model for Infant Treatment) model, which yields personalized length of stay estimates for an infant, utilizing data available from time of admission to the NICU.**

I. Introduction

It has been well established that the nurse to patient ratio has a direct relationship to patient morbidity [1], [2]. In fact earlier studies have shown that this relationship not only extends from general hospital care to the nurse-infant ratio in the NICU, but may actually become more pronounced in intensive care units such as these [3]–[5]. Beyond the obvious danger of inadequate nurse ratios on infant care, understaffing presents an additional risk through the inclusion of float nurses. A float nurse is an attempt to adjust for staffing issues in a hospital, utilizing nurses from a similar unit that is overstaffed to work in the at risk unit. For the NICU nurses typically float from Pediatrics or the Pediatric Intensive Care Unit. However, the practice of floating has not been well documented and may present additional issues due to a lack of familiarity with the patients and procedures of the new unit [6].

Further, unit staffing is not only a critical factor for infant outcomes, but also for the nurses themselves. The nature of care in the NICU has been shown to be highly stressful, and has been shown to be exacerbated in units that experience continual understaffing [7], [8]. It is important to note that prolonged stress at this level has been linked to increased job dissatisfaction and nurse "burn out syndrome" [2], [7].

It is then unsurprising that there has been a substantial amount of prior research into determining appropriate NICU staffing levels. Many traditional staffing techniques rely on features such as historical census count, while newer models may factor in the architectural features of a particular NICU. The increased focus on a patient centric atmosphere and improved family privacy, has led to the architectural redesign of many NICU units to utilize individual patient care rooms, colloquially called PODs. Recommendations for each POD designate a certain number of nurses as a base staffing required for the unit to operate safely [9], [10]. This then raises the question, *why then are staffing issues so ubiquitous in the NICU?* We believe the answer to this question stems from the unique care requirements of each infant, and can be contributed to the failure of traditional staffing models to accurately address all aspects of the infant's condition.

Unfortunately both architectural and traditional models fail to capture the critical acuity aspect of an infant's condition. Acuity is a proxy for the level of care each infant requires based on their condition, and the official staffing ratio guidelines published in [11] are based in part on this acuity level. For example two NICU may each have 20 admitted infants, but NICU-1 may contain all infants classified as "continuing care" which require a 1:4 nurse to infant ratio, while NICU-2 may have all "unstable" infants each requiring a 1:1 staffing ratio. It is clear that no historical model can account for this degree of granularity, and the problem is once again reduced to a reactive staffing exercise. Newer more complex models have been created to account for the volatile nature of the infant's conditions. However, as we will see these models suffer from a period of training lag, making them unusable for anywhere from 8-30% of the infants admitted to the NICU.

This paper will present a novel length of stay prediction model dubbed ADMIT (Admission Duration Model for Infant Treatment). The model is intended to allow charge nurses to accurately predict a personalized length of stay for each infant in the NICU. Allowing for significantly simplified staffing as the required nurse to infant ratio can be determined for extended periods of time, based on the empirical acuity of the infant and their modeled length of stay. Further unlike existing models that as we will see suffer from a training lag, the ADMIT model is intended to provide these predictions utilizing information available at the time of the infants admission.

II. Related Work

While there are currently no prior works focused directly on utilizing the length of stay for effective unit staffing, there are works which attempt to predict an infant's length of stay in a NICU based on data from electronic health records. These works present models which employ a variety of machine learning techniques such as artificial neural networks, linear and regression mixture models. Each utilizes various clinical

978-1-4799-5669-2/14/$31.00 ©2014 IEEE

features such as provider orders, procedures and medications to produce their predictions [12]–[14].

The issue with these prior works comes not based on their results, but rather from the types and quantity of information needed by each model. A closer look at the features utilized reveals many involve post-admission information, such as orders placed and procedures preformed. These features are a result of patient care, and must be generated over a period of time. This period forces a lapse from the time an infant is admitted to the NICU, until the point when the models have sufficient data to provide accurate length of stay predictions. However our data reveals 30% of all infants were discharged within 1 week. Thus this lapse then prevents the usage of these models in short term scheduling.

Further, long before ADMIT or any of the other models discussed above, hospitals had created their own prediction techniques. Two of the most well known are the CRIB, and SNAP models [15], [16]. The CRIB model focuses on features such as gestational age, and FiO_2, whereas the SNAP model utilizes 26 physiological features such as blood pressure, heart and respiratory rates. It is important to note that these models do not directly predict length of stay, but rather produce risk scores. However, it has been shown by Berry et al. that these scores can be used as a predictor of the length of stay [17].

In this case an issue arises in that while there may be a correlation between these scores and the infant's length of stay, these models were not created with the intention of being length of stay predictors. They both perform well for the task of predicting high-risk infants, but will not obtain the level of accuracy possible from a specialized model. Additionally both models make use of data obtained after admission, again subjecting them to the lag issue described above.

III. DATA

The data used for this analysis was provided as a part of the Vermont Uniform Hospital Discharge Data Set (VUHDDS) [18][1]. The VUHDDS is a dataset containing inpatient, emergency room and outpatient records of all fourteen of Vermont's general acute care hospitals. Each instance in the dataset represents a single patient and contains 77 features detailing all facets of the patient's hospitalization. These features include method of admission, hospital identifier, up to 19 procedure and 19 diagnosis codes including the required diagnosis related group (DRG) and major diagnostic category (MDC). The diagnoses provided by the VUHDDS are represented as International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes [19]. Further, the records also contain detailed socioeconomic information, such as the patient's home zip code, binned age range, gender and the principal payment type of their insurance.

[1]Hospital discharge data for use in this study were supplied by the Vermont Association of Hospitals and Health Systems-Network Services Organization (VAHHS-NSO) and the Vermont Department of Financial Regulation (DFR). All analyses, interpretations or conclusions based on these data are solely that of the authors. VAHHS-NSO and DFR disclaim responsibility for any such analyses, interpretations or conclusions. In addition, as the data have been edited and processed by VAHHS-NSO, DFR assumes no responsibility for errors in the data due to coding or processing by hospitals, VAHHS-NSO or any other organization, including the authors

Feature	Code	Value
Admission Source	15	Newborn Born in this Hospital
Patient Home Zip Code	05401	Burlington
Gender	1	Male
Insurance Principle Payer	6	Blue Cross
Second Diagnosis	7746	Fetal/neonatal jaund NOS
Hospital Healthcare Service Area	2	Burlington
Critical Access Hospital Flag	0	Not a Critical Access Hospital
SCU Days - *Class*	12	Days in a Special Care Unit

TABLE I: Sample Infant Instance

A. Feature Selection

As this work is focused on the determination of length of stay from the time of *admission* many of the 77 features were removed. Primarily these were related to the patient's discharge including discharge status, and the DRG and MDC codes, as these codes are calculated only after all diagnosis and procedural information have been collected. Additionally, as the emergency, procedure and diagnosis codes were not provided with timestamp information it could not be determined if any were collected prior to the infant's admission. As such all were removed from consideration except for the secondary diagnosis, justified below. In the end seven features, and one class value remained, a sample instance can be seen in Table I. Of the final feature vector 4 of the 7 features pertained to either the socioeconomic data of either the patient or hospital itself. This is logical, as this information is the most readily available at the time of admission. Of the seven features, two in particular warrant additional consideration.

The Critical Access Flag: Indicates a "limited-service hospital designed to provide essential services to rural and frontier communities" [20]. As such these hospitals are likely not equipped to handle the sickest and smallest newborns. The infants who need such a level of care are likely born at, or transported to, a more equipped NICU. For the sake of this work we assume that the infant is provided the appropriate level of care at the hospital indicated.

Secondary Diagnosis: While it may appear that we are violating the claim that the ADMIT model works at the time of admission by using the *secondary diagnosis*, we claim that this is in fact the diagnosis with which an infant would be admitted into the NICU as the primary diagnosis is typically related to their type of birth itself, see Table II.

While these codes may be useful for an insurance company, they are not diagnostically very useful to a NICU. Infants are admitted directly from labor and delivery or from another NICU through transfer, and would typically be admitted with a specific diagnosis. As the primary diagnosis will contain the birth details, the secondary diagnosis must then contain the admission diagnosis. In fact the primary diagnosis had only 51 unique values, while the secondary diagnosis presented 525 unique values. For comparison, Table III details the top 3 most prevalent secondary diagnoses. It is important to note here that N/a is a legitimate secondary diagnosis, likely indicating no additional complications to the birth.

B. Infant Data

The infant data was obtained by filtering inpatient records whose admission *type* was marked as "newborn"; it should be noted that the admission *source* might vary based on whether

Primary Diagnosis	Value
V3000	Single Liveborn Born In Hospital, Delivered Without Mention Of Cesarean Section
V3001	Single Liveborn Born In Hospital, Delivered By Cesarean Section
V3101	Twin Birth Male Liveborn, Born In Hospital, Delivered By Cesarean Section

TABLE II: Top 3 most prevalent primary diagnosis

Secondary Diagnosis	Value
N/a	No Secondary Diagnosis
7746	Unspecified fetal and neonatal jaundice
76621	Post-term infant

TABLE III: Top 3 most prevalent secondary diagnosis

the infant was born in-hospital or transferred from another NICU. For this work a total of 28,087 infant records were aggregated over 5 years of records from 2008 through 2012. As with all data collected from actual working environments there were some quality issues. Detailed examination highlighted 26 records with ages in the 20's or 30's. This is clearly an error, and these records were subsequently removed, resulting in total of 28,061 records used for the remainder of this work. It is important to note that the VUHDDS does not directly specify days spent in an NICU, rather the dataset provides two distinct length of stay metrics, one for total patient days and one for days spent in Special Care Units (SCU). However we claim the only unit equipped to treat those patients marked as "newborn" would be the NICU. Thus the SCU days can be considered a direct proxy for the days spent in the NICU for each infant.

C. Data Analysis

Critical Care by nature is a heavily imbalanced problem, with 96.05% of the 28,061 newborns in the dataset spending 0 days in a special care unit (NICU). As stated in prior works, the length of stay statistics for the remaining 1,108 infants who were admitted to the NICU are heavily right skewed. Looking further we can see 8.39% of all infants admitted to the NICU leave after only a single day under care, this is most likely due to those infants who are admitted for distress during delivery, but after observation have no adverse effects.

Additionally, while 1,108 infants may appear to be a small dataset we can show that this figure is indeed reflective of the true admission rate. These 1,108 infants came from a total of 28,061 births over 5 years provided by the dataset, averaging to 5,612.2 births per year captured by the VUHDDS. As a point of reference in the year 2012 the CDC recorded a total of 6,009 births in the state of Vermont[2]. This shows that the VUHDDS average captured approximately 93.40% of all births in the state. Additionally as the number of births continues to grow annually the VUHDDS likely captures a slightly higher percent of total births on a year-by-year basis. Thus, while the figure of admitted infants seems low we have no reason to believe the imbalance is not reflective of the true percentage of admissions to a NICU.

IV. Initial Analysis

We began our evaluation by considering the number of days admitted to the NICU as a continuous function. Beyond standard regression classifiers we utilized models specifically created for datasets containing low occurrence high importance instances, due to the admission length imbalance. The results from the regression evaluations can be seen in Table VI.

Several of the models performed well demonstrated not only by their high correlation coefficient to the true length of stay, but also in their Mean Absolute Error (MAE). The MAE is a measure of closeness to the true length of stay, and thus we can see these models consistently produce predictions within 1 day of the infants true stay duration. However, despite their demonstrated accuracy, all of the evaluated regression models produced significant variance within their length of stay predictions. The Root Mean Square Error (RMSE) is representative of a models bias-variance tradeoff, and when used in conjunction with the MAE can be used to discern a sense of magnitude for the variance. As seen in Table VI the RMSE is substantially larger then the MAE, which indicates that the result includes a large variance[3]. As with MAE, RMSE is provided in the same units as the length of stay, indicating a variance on the order of multiple days. This range of variance effectively renders these models useless in a clinical setting.

As a result it became clear that we required a fundamental shift in our consideration of the length of stay problem. Prior work by Torgo et al. has detailed the use of set intervals to transform a regression problem into discrete classes on which to learn a classification model [21]. However, while they transform the data back before evaluation, we maintain the discretized data for evaluation step. We began with the most fundamental transformation, treating each day as a discrete class. Not only does this involve the most direct transformation process, it also yields the most detailed estimate of the number of days an infant will spend in a NICU.

V. Model Design

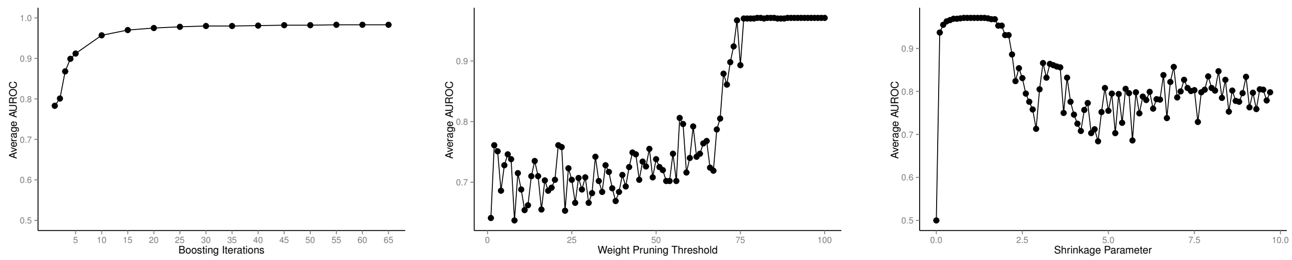
The ADMIT model utilizes an augmentation of the Ada-Boost algorithm, known as LogitBoost [22]. It is trained on all 28,061 infants, using the reduced 7-feature dataset. The decision to utilize LogitBoost over the standard Ada-Boost algorithm was based on consideration of each algorithm's loss function [23]. While similar, an added logistic term in LogitBoost performs a "dampening" effect during the feature reweighing. As boosting ensembles are designed to utilize weak classifiers in an iterative fashion, at each iteration the training instances are dynamically reweighted based on the incorrect predictions from the prior iteration. Due to the imbalanced nature of the length of stay problem it is likely that the base classifier will initially perform poorly. However without the logistic dampening, the standard Ada-Boost has been shown to potentially reweight these instances too aggressively, actually decreasing the overall performance [24], [25].

The base learner utilized by the ADMIT model is the decision stump, as the branching step does not rely on any particular distance measure. This becomes particularly important when utilizing features such as ICD9-CM codes, which have no inherent concept of distance. While it is true these codes are grouped by related condition, codes 460-519 (diseases of the respiratory system[4]) are not more or less

[2] http://www.cdc.gov/nchs/data/nvsr/nvsr62/nvsr62_09.pdf

[3] http://www.eumetcal.org Mean Absolute Error & Root Mean Squared Error

[4] http://www.health.gov.bc.ca/msp/infoprac/diagcodes/



(a) Boosting Iterations (b) Weight Threshold (c) Shrinkage

Fig. 1: ADMIT Model - Parameter Tuning

similar to those groups in either direction. ICD-CM codes 390-459 represent diseases of the circulatory system, while codes 520-579 represent diseases of the digestive system.

Finally unlike more complex algorithms such as Support Vector Machines (SVM), LogitBoost has only three main parameters that directly effect performance. These are the number of boosting iterations, weight threshold and shrinkage parameter. Optimal values of 30, 100 and 1 respectively were confirmed through evaluation of AUROC, seen in Figure 1.

VI. Results and Analysis

It is important to note that 5-fold cross-validation was used for all evaluations. The decision to utilize 5, over 10-fold was a result of the data imbalance. As the majority of infants are not admitted to the NICU, increasing the fold count and thus decreasing the hold-outset size, would result in each fold containing a larger percentage of non-admitted infants and thus skewing the results [26].

A. *Classification Models*

In addition to the evaluation of the ADMIT model we compared our analysis against a representative set of classification models, representing three main categories. Firstly we included common baseline models such as Naive Bayes and C4.5. Secondly, models which focus on non-linearly separable data such as SVM. Finally as with the regression models, those that have been proven effective on imbalanced and noisy data such as Ada-Boost and Random Forrest. The results from all classification model evaluations can be seen in Table VII.

Analysis: An initial observation was the substantially reduced MAE and RMSE, indicating a much improved error bound over the prior regression models. Looking further, the effectiveness of Random Forest and Naive Bayes indicated some degree of noise within the dataset, but interestingly the typically high performing SVM produced results equivalent to a random classifier. This may be a result of SVM's reliance on a Euclidian distance metrics, which have been shown to perform sub-optimally on imbalanced data [27]. However, with the distance invariant decision stump as the base learner it was surprising that Ada-Boost performed so poorly. As mentioned above, literature suggests this it could be a result of an overfitted model due to aggressive instance reweighting. This is supported by the overall performance of the ADMIT model found at the bottom of Table VII.

Class	**ROC Area**	**Class**	**ROC Area**
0-7.071	0.980	49.500-56.571	0.965
7.071-14.142	0.947	56.571-63.643	0.946
14.142-21.214	0.958	63.643-70.714	0.931
21.214-28.286	0.961	70.714-77.786	0.867
28.286-35.357	0.962	77.786-84.857	0.951
35.357-42.429	0.981	84.857-91.929	0.759
42.429-49.50	0.954	91.929-99	0.939
Avg. AUROC		*0.979*	

TABLE IV: Weekly Discharge Performance - ADMIT

Feature	Correlation Value
Critical Access Hospital Flag	0.106
Second Diagnosis	0.053
Healthcare Service Area	0.038
Zip Code	0.026
Sex	0.012
Principle Payer	0.010
Admission Source	0.007

TABLE V: Pearson's Correlation Results

Our next step was to demonstrate the generalizability of the ADMIT model by evaluating the performance over larger class intervals. We created another transformation increasing the interval size from a single day to represent weekly discharges, which resulted in 14 distinct classes. The results for each class are detailed in Table IV. Finally, we conducted a correlation analysis to ensure that none of the features contained latent information correlated to the length of stay in the NICU. The analysis utilized the Pearson's correlation, and the results can be seen in Table V. With a value of only 0.106 even the highest correlation fails to indicate any significant influence over the number of days an infant will spend in the NICU, again demonstrating the effectiveness of the ADMIT model.

VII. Conclusion

To recap, this work provides a tool to aid in the dynamic requirements of NICU nurse staffing. Used in conjunction with an infant's acuity level the ADMIT model will allow charge nurses to gain a deeper understanding of the future staffing requirements, utilizing each infant's personalized expected length of stay. The results show the ADMIT model performs well, and has been evaluated on a representative dataset, spanning 5 years and incorporating 14 hospitals located in 13 different locations within Vermont. This model will hopefully aid in furthering the overarching goal of improved quality of care and patient safety.

	Model	Correlation Coefficient	Mean Absolute Error	Root Mean Squared Error
	Linear Regression	0.712	1.008	4.036
Regression	Additive Regression	0.683	0.904	4.195
	Reduced Error Pruning	0.719	0.729	4.019
	M5- Regression Tree	0.723	0.754	3.966
Tree Based Models	M5- Model Tree	0.744	0.716	3.847
Nearest Neighbour	KNN	0.5318	0.862	4.957
Rule Based Models	M5-Rule Learner	0.746	0.715	3.835

TABLE VI: Regression Model Performance

	Model	Average AUROC	Mean Absolute Error	Root Mean Squared Error
Naive Bayes	Naive Bayes	0.923	0.001	0.026
	SVM - Radial Kernel	0.500	0.020	0.099
Support Vector Machine	SVM - Polynominal Kernel	0.577	0.020	0.020
	Reduced Error Pruning	0.499	0.001	0.0268
Tree Based Models	C4.5	0.499	0.001	0.027
Nearest Neighbour	KNN	0.911	0.001	0.027
	Ripper	0.514	0.001	0.027
Rule Based Models	PART	0.932	0.001	0.025
	Ada-Boost - Decision Stump	0.586	0.016	0.080
Ensemble Methods	Random Forrest 50 Trees	0.939	0.001	0.026
	ADMIT		*0.971*	

TABLE VII: Classification Model Performance

REFERENCES

[1] J. Needleman, P. Buerhaus, V. S. Pankratz, C. L. Leibson, S. R. Stevens, and M. Harris, "Nurse staffing and inpatient hospital mortality," *New England Journal of Medicine*, vol. 364, no. 11, pp. 1037–1045, 2011.

[2] L. H. Aiken, S. P. Clarke, D. M. Sloane, J. Sochalski, and J. H. Silber, "Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction," *Jama*, vol. 288, no. 16, pp. 1987–1993, 2002.

[3] L. Callaghan, D. Cartwright, P. ORourke, and M. Davies, "Infant to staff ratios and risk of mortality in very low birthweight infants," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 88, no. 2, pp. F94–F97, 2003.

[4] K. E. S. Hamilton, M. E. Redshaw, and W. Tarnow-Mordi, "Nurse staffing in relation to risk-adjusted mortality in neonatal care," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 92, no. 2, pp. F99–F103, 2007.

[5] R. L. Kane, T. A. Shamliyan, C. Mueller, S. Duval, and T. J. Wilt, "The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis," *Medical care*, vol. 45, no. 12, pp. 1195–1204, 2007.

[6] J. Dziuba-Ellis, "Float pools and resource teams: A review of the literature," *Journal of nursing care quality*, vol. 21, no. 4, pp. 352–359, 2006.

[7] C. Walker, "Neonatal intensive care and stress." *Archives of disease in childhood*, vol. 57, no. 2, p. 85, 1982.

[8] S. P. Jacobson, "Stressful situations for neonatal intensive care nurses," *MCN: The American Journal of Maternal/Child Nursing*, vol. 3, no. 3, pp. 144–150, 1978.

[9] R. White, J. Smith, and M. Shepley, "Recommended standards for newborn icu design," *Journal of Perinatology*, vol. 33, pp. S2–S16, 2013.

[10] D. R. Thompson, D. K. Hamilton, C. D. Cadenhead, S. M. Swoboda, S. M. Schwindel, D. C. Anderson, E. V. Schmitz, A. C. S. Andre, D. C. Axon, J. W. Harrell *et al.*, "Guidelines for intensive care unit design*," *Critical care medicine*, vol. 40, no. 5, pp. 1586–1600, 2012.

[11] C. J. Lockwood, J. A. Lemons, A. A. of Pediatrics, A. C. of Obstetricians, Gynecologists *et al.*, *Guidelines for perinatal care*. American Academy of Pediatrics, 2007.

[12] B. Zernikow, K. Holtmannspötter, E. Michel, F. Hornschuh, K. Groote, and K.-H. Hennecke, "Predicting length-of-stay in preterm neonates," *European journal of pediatrics*, vol. 158, no. 1, pp. 59–62, 1999.

[13] K. K. Yau, A. H. Lee, and A. S. Ng, "Finite mixture regression model with random effects: application to neonatal hospital length of stay," *Computational statistics & data analysis*, vol. 41, no. 3, pp. 359–366, 2003.

[14] S. R. Levin, E. T. Harley, J. C. Fackler, C. U. Lehmann, J. W. Custer, D. France, and S. L. Zeger, "Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders," *Critical care medicine*, vol. 40, no. 11, pp. 3058–3064, 2012.

[15] T. I. N. Network, "The crib (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units," *The lancet*, vol. 342, no. 8865, pp. 193–198, 1993.

[16] D. K. Richardson, J. E. Gray, M. C. McCormick, K. Workman, and D. A. Goldmann, "Score for neonatal acute physiology: a physiologic severity index for neonatal intensive care," *Pediatrics*, vol. 91, no. 3, pp. 617–623, 1993.

[17] M. Berry, P. Shah, R. Brouillette, and J. Hellmann, "Predictors of mortality and length of stay for neonates admitted to children's hospital neonatal intensive care units," *Journal of Perinatology*, vol. 28, no. 4, pp. 297–302, 2008.

[18] "Vermont uniform hospital discharge data set," http://healthvermont.gov/research/hospital-utilization.aspx.

[19] V. N. Slee, "The international classification of diseases: ninth revision (icd-9)," *Annals of internal medicine*, vol. 88, no. 3, pp. 424–426, 1978.

[20] J. F. Ahrens, "Critical access hospitals: Their role and policy implications in primary and long-term care," The Maureen and Mike Mansfield Foundation, Tech. Rep., the Maureen and Mike Mansfield Foundation.

[21] L. Torgo and J. Gama, "Regression by classification," in *Advances in artificial intelligence*. Springer, 1996, pp. 51–60.

[22] J. Friedman, T. Hastie, and R. Tibshirani, "Special invited paper. additive logistic regression: A statistical view of boosting," *Annals of statistics*, pp. 337–374, 2000.

[23] R. E. Schapire, "The boosting approach to machine learning: An overview," in *Nonlinear estimation and classification*. Springer, 2003, pp. 149–171.

[24] M. Dettling and P. Bühlmann, "Boosting for tumor classification with gene expression data," *Bioinformatics*, vol. 19, no. 9, pp. 1061–1069, 2003.

[25] K.-Y. Feng, Y.-D. Cai, and K.-C. Chou, "Boosting classifier for predicting protein domain structural class," *Biochemical and Biophysical Research Communications*, vol. 334, no. 1, pp. 213–217, 2005.

[26] T. Raeder, G. Forman, and N. V. Chawla, "Learning from imbalanced data: evaluation matters," in *Data Mining: Foundations and Intelligent Paradigms*. Springer, 2012, pp. 315–331.

[27] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani, *The elements of statistical learning*. Springer, 2009, vol. 2, no. 1.