

# Predicting Neonatal Encephalopathy From Maternal Data in Electronic Medical Records

Thomas Li<sup>1</sup>, Cheng Gao, PhD<sup>2</sup>, Chao Yan, MS<sup>2</sup>, Sarah Osmundson, MD<sup>2</sup>, Bradley A. Malin, PhD<sup>2</sup>, You Chen, PhD<sup>2</sup>

<sup>1</sup>Duke University, Durham, NC; <sup>2</sup>Vanderbilt University, Nashville, TN

## Abstract

*Neonatal encephalopathy (NE) is a leading cause of neonatal mortality and lifetime neurological disability. The earlier the risk of NE can be assessed, the more effective interventions can be in preventing adverse outcomes. Existing studies that focus on intrapartum risk factors do not provide the early prognostic forecasting necessary to prepare healthcare professionals to intervene early in a high-risk NE case. This work used maternal data in a supervised machine learning framework to predict NE events. Specifically, we 1) collected the electronic medical records (EMRs) for 104 NE newborns and 31,054 non-NE newborns and their mothers, 2) trained and tested a regularized logistic regression on imbalanced and high-dimensional EMR data, and 3) discerned important features that could be possible risk factors. The learned model offers prenatal predictions of NE cases with an average area under the receiving operator characteristic curve (AUC) of 87% and identified the most important predictors.*

## 1 Introduction

Neonatal encephalopathy (NE) is a leading cause of infant mortality and long-term neurological morbidity<sup>1,2</sup>. The American College of Obstetricians and Gynecologists (ACOG) defines it as a syndrome of disturbed neurological function in the earliest days of life in an infant born at or beyond 35 weeks of gestation, manifested by a subnormal level of consciousness or seizures, and often accompanied by difficulty with initiating and maintaining respiration and depression of tone and reflexes<sup>1</sup>. Historically, the cause of NE was assumed to be asphyxial in nature, and the term hypoxic-ischemic encephalopathy was widely used<sup>2</sup> until the broader term neonatal encephalopathy was recently standardized by the ACOG. The change in terminology demonstrates a paradigm shift and recognition that the syndrome has multiple causal pathways. It was found that 70% of NE cases were most likely attributed to antepartum risk factors like abnormal fetal growth, maternal infection, and socioeconomic status<sup>3,4</sup>. Recognizing distal risk factors such as these are important because they may exert early pathological damage and, as an artifact, predispose the fetus to NE. If this is the case, then early, preventative interventions can minimize the effects of distal risk factors and decrease the risk of NE<sup>1</sup>.

Traditionally, approaches based on the knowledge of clinical experts have been relied upon to learn risk factors associated with NE. Domain expertise is leveraged to determine a set of candidate variables, after which statistical models are applied to ascertain which variables demonstrated an association with NE<sup>3-10</sup>. Most notably, a Western Australian study found high birth weight to be highly correlated with NE through a univariate analysis<sup>3</sup>. Another study considered over 50 antepartum and intrapartum features and constructed a high performance logistic regression model using eight features that were identified as significantly associated with NE via univariate chi-squared tests<sup>10</sup>. There are, however, several limitations to these previous approaches. First, this type of investigation restricts the study to a small number of variables, making it difficult to identify novel risk factors. Second, since these approaches rely on the expert knowledge, the variables investigated may be subject to selection bias, such that potential predictor variables could be neglected. Third, previous studies focused on assessing the risk of NE only after birth, using variables that often are inaccessible before delivery, such as birth weight and Apgar scores.

To the best of our knowledge, there has not been an investigation into the extent to which NE can be predicted ahead of birth. If preterm predictions could be made, clinicians could be afforded an opportunity to improve the quality of NE care or prevent it from developing. For instance, it has been shown that planned or elective Cesarean delivery is associated with a reduced incidence of infants suffering from NE<sup>12-14</sup>. To this end, we aimed to leverage antenatal maternal variables to model the risk of NE before birth. Maternal features have been extensively studied<sup>3,5,7,15,16</sup>. However, maternal features were often treated as discrete variables and mixed in with intrapartum and neonatal features to do NE risk assessment. In the context of NE, there has not been a targeted effort to diversely investigate the

**Table 1:** Summary of the VUMC Maternal Dataset. (W = White; B = Black; O = Other)

#Mother-Infant Pairs	Mother Age	Mother Race	#NE Cases	#Controls	#Unique ICD-9	#Unique CPT
31,158	23 [12,50]	W: 66%, B: 17.2%, O: 16.8%	104	31,054	7,860	6,095

maternal profile as a whole, perhaps because expert-driven approaches cannot investigate a large number of features, and consequently, cannot capture the full diversity of the maternal condition during pregnancy.

Our investigation was motivated by the need for a data-driven approach that models the risk of NE from a broad set of maternal variables. Our goal was to study the predictor variables that can present early in the NE causal pathway while mitigating variable selection bias. To this end, we present a retrospective study of 31,158 mother-infant pairs from over a decade of maternal electronic medical record (EMR) data at Vanderbilt University Medical Center (VUMC)<sup>17</sup>. The maternal variables were defined by billing codes that represent symptoms, diagnosis, and procedures, as well as demographic information. Given the low incidence of NE, we adopt a bootstrapping strategy to handle the imbalanced data (104 NE cases vs. 31,054 NE controls). We introduce a regularized logistic regression model that ranks the importance of maternal variables in terms of their ability to distinguish NE cases from control.

## 2 Materials and Methods

### 2.1 Materials

#### 2.1.1 Dataset

The dataset was drawn from the EMR system at VUMC, which covers over 2.5 million patient records over a 29-year period. All data were de-identified and the study was approved by the Vanderbilt IRB. Table 1 summarizes various features about the population and the balance of the dataset. We identified 31,158 links between mothers and infants. The EMRs of the infants were inspected for the purpose of case identification of NE<sup>1</sup>. Maternal EMR information prior to delivery was collected to train and test the classification models, as well as the subsequent analysis of risk factors. Specifically, for each delivery event, we collected age, race, ICD-9 and CPT codes assigned before the time of the linked delivery. As shown in Table 1, the average age of mothers at delivery was 23 years old, with the youngest at age 12 and the oldest at age 50. The mothers were 66% White, 17.2% Black and 16.8% some other races. There were 7,860 unique ICD-9 codes and 6,095 unique CPT procedural codes observed in the mothers' EMRs prior to delivery.

#### 2.1.2 Identification of Cases and Controls

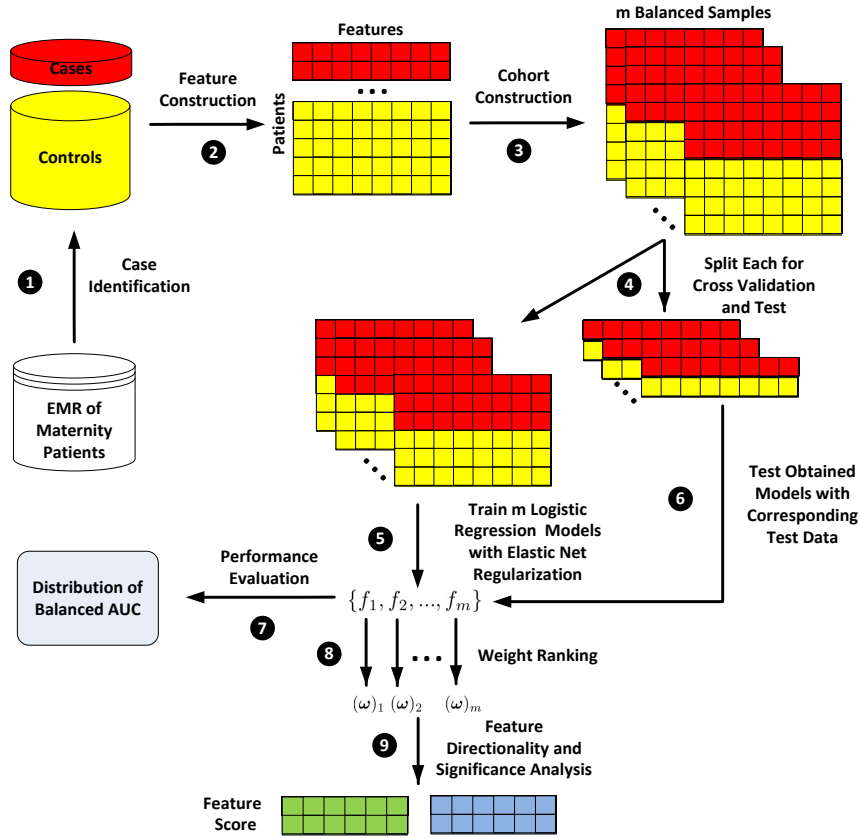
In this study, a *case* corresponds to a mother who had at least one infant who was diagnosed with NE. We identified cases through a two-step process. First, we automatically identified NE candidates from the clinical notes of the delivered infants of the mothers by matching keywords in the natural language. Specifically, we conducted a free text search using regular expressions built on a set of keywords: {HIE, hypoxic-ischemic encephalopathy, hypoxemia, neonatal encephalopathy, asphyxia}. In doing so, we extracted 141 NE candidates. Since matching a keyword does not guarantee the actual incidence of the disease, in the second step, we manually reviewed each of the clinical notes of NE candidates, and with the confirmation of a clinically-knowledgeable expert, we labeled 104 of the candidates as NE and their mothers as cases. One of the primary reasons that the 37 candidates were excluded is that their clinical notes had some expression of a negation of the matched term (e.g., the infant did not suffer from HIE).

We defined a *control* as a mother who did not deliver any infants diagnosed with NE during a birthing event. As such, controls include mothers whose infants did not match on any term in the keyword list, as well as the mothers of infants who were considered as NE candidates in the first step of case identification but filtered by manual review at the second step. In the end, the number of controls was 31,054. Note that the case identification is based on each delivery event, such that a mother who gave birth to both a NE infant and a non-NE infant(s) and could be labeled as both a case and control (though this would constitute two separate instances). However, we did not observe any mothers that would be labeled as both a case and control.

<sup>1</sup>The VUMC creates an EMR for each delivered infant.

## 2.2 Methods

We designed a data-driven framework that consists of 5 core modules: 1) feature construction to characterize each case/control, 2) cohort establishment to handle the imbalanced data, 3) predictive model training, 4) performance evaluation to assess predictive capability, and 5) feature analysis to determine which features were most associated with NE. The 5 modules are depicted in Figure 1. Details for each module are reported in the following subsections.



**Figure 1:** A framework for NE prediction and risk factor analysis based on EMR data.

### 2.2.1 Feature Construction

A feature space from the records of mothers was constructed. For each mother in the cases/controls, we extracted age, race, and the ICD-9 and CPT codes that were documented in the EMR within one year prior to delivery. The billing codes were limited to one year prior to delivery in an attempt to control for variability of observational window duration between cases and controls. This decision is explained in more depth in the discussion section. To reduce the dimensionality of the feature space and mitigate redundancy, we generalized ICD-9 codes by rolling them up the hierarchy as follows. First, all codes that pertain to pregnancy, childbirth, or conditions originating from the perinatal period (e.g., 630.\* to 679.\*, 760.\* to 779.\*, and 23.\* to 23.9\*) were rolled up to one character after their decimal. Second, all other codes were rolled up to the set of characters before the decimal point (e.g., 320.1 and 320.82 both generalized to 320).

As summarized in Table 2, all of the ICD-9 and CPT codes were represented as binary features, indicating the presence or absence of such codes in the EMR. Age and race were represented as ordinal and nominal features respectively. We borrowed the age categories from previous clinical studies on NE<sup>3,5</sup>. Since only a small number of mothers fell into

**Table 2: Summary of Features**

Feature Type	ICD-9 codes	CPT codes	Age			Race		
Value	{0,1}	{0,1}	1: [20, 29],	2: [30, 34],	3: greater than 34	1: White,	2: Black,	3: Other

race categories other than “black” or “white”, we grouped them into the “other” category.

### 2.2.2 Cohort Construction

As mentioned earlier, our dataset is highly imbalanced in terms of cases and controls, which is a concern for several reasons. First, it is widely accepted that learning from an imbalanced dataset leads to models that lack stability and exhibit poor generalization performance<sup>18</sup>. Second, there is a strong possibility that training in this setting will yield a model that simply predicts each newborn as a control (i.e., no NE). The model would achieve both a high accuracy and high false negative rate (i.e., predicting cases as controls) and would, thus, lack clinical utility. To address the class imbalance problem, we adopted an undersampling technique to construct a set of controls that contain the same number of instances as the cases.

Given that a single sample, although constructed randomly, may yield a biased prediction model, we conducted sampling in a bootstrapped manner, such that we composed  $k$  sample sets to evaluate the variance and confidence interval of the learned statistics. Specifically, we sampled 300 sets of 104 controls, each of which was analyzed with respect to the same set of 104 cases. For each cohort, features that appeared in less than 1% of the instances in a single cohort were eliminated from that cohort’s feature space to reduce sparsity and streamline data transfer.

### 2.2.3 Predictive Models

To train the models and evaluate their performance on our 300 cohorts, we used 75% of the data to perform a stratified sampling within each cohort for 10-fold cross-validation. The remaining 25% of the cohort served as a test dataset. This sampling strategy ensures that the models are both trained and evaluated on the rebalanced data.

During the training phase, we trained a regularized logistic regression model with a mixture of the  $\ell_1$  and  $\ell_2$  penalties, also known as the elastic net penalty, that identified important features. This method has been shown to perform better at various classification problems than both the  $\ell_1$  or  $\ell_2$  penalties alone on input spaces where features greatly outnumber the number of samples. Regularization is designed to limit error due to variance, and thus improve the generalizability of a trained model<sup>19</sup>. This is a highly desirable quality for our use case of training many models on many cohorts from a single population because all the models should perform consistently well, and be robust to the noise in the specific cohort they were trained on.

Regularized logistic regression is able to perform automatic feature selection<sup>19,20</sup> by ranking features and selecting a subset of heavily weighted features. In doing so, it eliminates or reduces the impact of less important features by setting their coefficients to zero or close to zero. The coefficients corresponding to features, i.e. the parameters of the model, indicate the importance of features. Based on this metric, feature selection is conducted.

After executing the training and validation module, we generate 300 predictive models with different coefficients, which we represent as  $\{f_1, f_2, \dots, f_m\}$  in Figure 1.

### 2.2.4 Performance Evaluation and Feature Ranking

To evaluate the generalization performance of each of the learned predictive models, we applied each model on its corresponding test dataset (i.e., a group of the remaining 25% samples in each cohort). We report the distribution of the area under the receiver operating characteristic curve (AUC) measure. Moreover, we record precision, sensitivity, and specificity for each model and computed the mean and the 95% confidence interval of AUC, precision, sensitivity and specificity values of all 300 models.

**Table 3:** Summary of Cohort Features

	# Sparse Features Eliminated	# ICD-9 codes	# CPT codes	# Selected Features
<b>Mean</b>	796.6	213.8	336.8	45.06
<b>95% Confidence Interval</b>	[782.1, 811.0]	[212.7, 214.9]	[335.1, 338.5]	[41.34, 48.78]

As mentioned above, regularization distinguishes important features<sup>2</sup> from among a large feature space in each classifier. These features are either positive or negative indicators for the presence of NE. To a certain degree, the learned coefficients reflect the relative contribution of the corresponding features to the NE prediction. If 0 represents the non-NE class and 1 represents the NE class, then the greater the weight of a certain feature, the more its presence relates to the NE. In contrast, the more negative a weight is, the more its presence is associated with a non-NE outcome. If a feature is eliminated from a model, then its corresponding coefficient will be zero.

To determine the predictive performance of a feature for NE, we computed the proportion that a feature’s weight covers with respect to the sum of all weights in the model. These proportions were then averaged across all models that utilized that feature, or, in other words, all models that retained the feature with a non-zero coefficient. We refer to this value as the feature *importance*, which is formalized in Equation 1, where the importance of feature  $x_p$  is computed based on the learned  $m \times n$  weights  $\{\omega_k^i\}$  from  $m$  classifiers and  $n$  features. We also designed another metric, *average frequency*, to measure the contribution of each feature to NE classification. Average frequency for feature  $x_p$  is calculated as the fraction of models having a non-zero coefficient for feature  $x_p$  out of all of the  $m$  models. This metric is formalized as Equation 2, where function  $I(\gamma)$  equals to 1 if condition  $\gamma$  is satisfied, 0 otherwise.

$$Importance(x_p) = \frac{1}{m} \sum_{i=1}^m \sum_{k=1}^n \omega_p^i / \omega_k^i \quad (1)$$

$$Average\_Frequency(x_p) = \frac{1}{m} \sum_{i=1}^m I(\omega_p^i \neq 0). \quad (2)$$

After computing the importance and the average frequency of each feature, we were able to rank features accordingly. Features above a certain rank are regarded as potentially important features, which, may serve as early evidence for either the presence or absence of NE.

### 3 Results

An average of 796.6 blank sparse features (occurring in less than 1% of examples within a cohort) were eliminated from each cohort. Table 3 shows that after feature and cohort construction steps, cohorts had 213.8 ICD-9 codes on average and 336.8 CPT codes on average, constituting 552.6 features on average including Age and Race. Models had an average of 45.06 features with non-zero coefficients. In other words, an average of 507.54 features were eliminated during model training. In the subsequent sections, we report the results of 1) the balanced performance of 300 predictive models on the NE prediction and 2) the importance of the features.

#### 3.1 The balanced AUC, precision, sensitivity and specificity

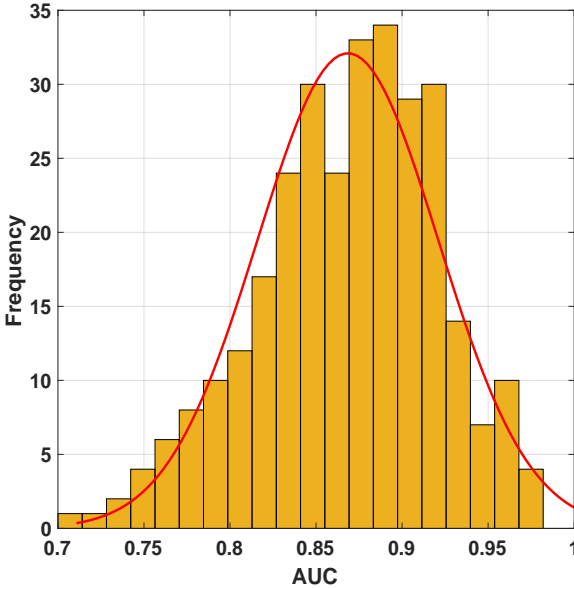
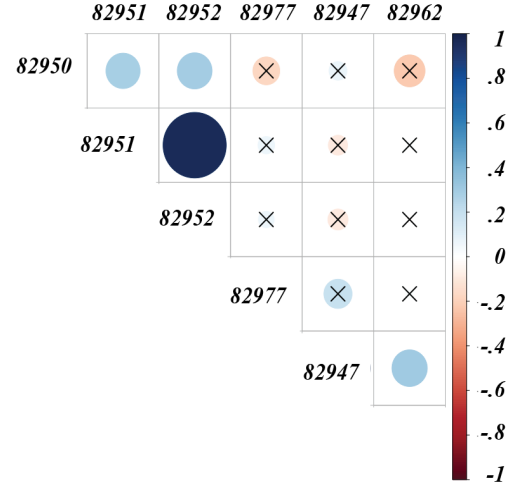
For each of the 300 predicted models, we recorded their AUC score, precision, sensitivity, and specificity. The distribution of the predictive models as a function of AUC is shown in Table 5. It can be seen that 89% of the models achieved an AUC larger than 0.8. The mean of the AUC scores is 0.869 and its distribution is shown in Figure 2.

A Kolmogorov-Smirnov Test<sup>21</sup> failed to reject a null hypothesis that the sample distribution of AUC scores satisfies a normal distribution at the 0.05 significance level. The 95% confidence interval for the mean of AUC scores was 0.8625 to 0.8745. The small confidence interval suggests that proposed framework can support NE prediction models with both stable and high predictive performance.

<sup>2</sup>The threshold  $|w| > 0$  is used to define the importance of features.

**Table 4:** Performance of the NE models.

	AUC	Precision	Sensitivity	Specificity
Mean	0.8685	0.8080	0.8079	0.8104
95% Confidence Interval	[0.8625, 0.8745]	[0.8000, 0.8159]	[0.7982, 0.8177]	[0.8014, 0.8194]
K-S Test	Normal	Normal	Normal	Normal

**Figure 2:** Histogram of AUC scores of 300 models. The normal distribution in red curve is generated based on the obtained mean and the standard deviation.**Figure 5:** A plot of the correlation matrix for CPT codes in the 829 group. Crosses indicate that the relationship between the corresponding pair of codes failed to be confirmed as significant. Blue and red represent positive and negative relationships respectively. The larger the circle the greater the correlation.

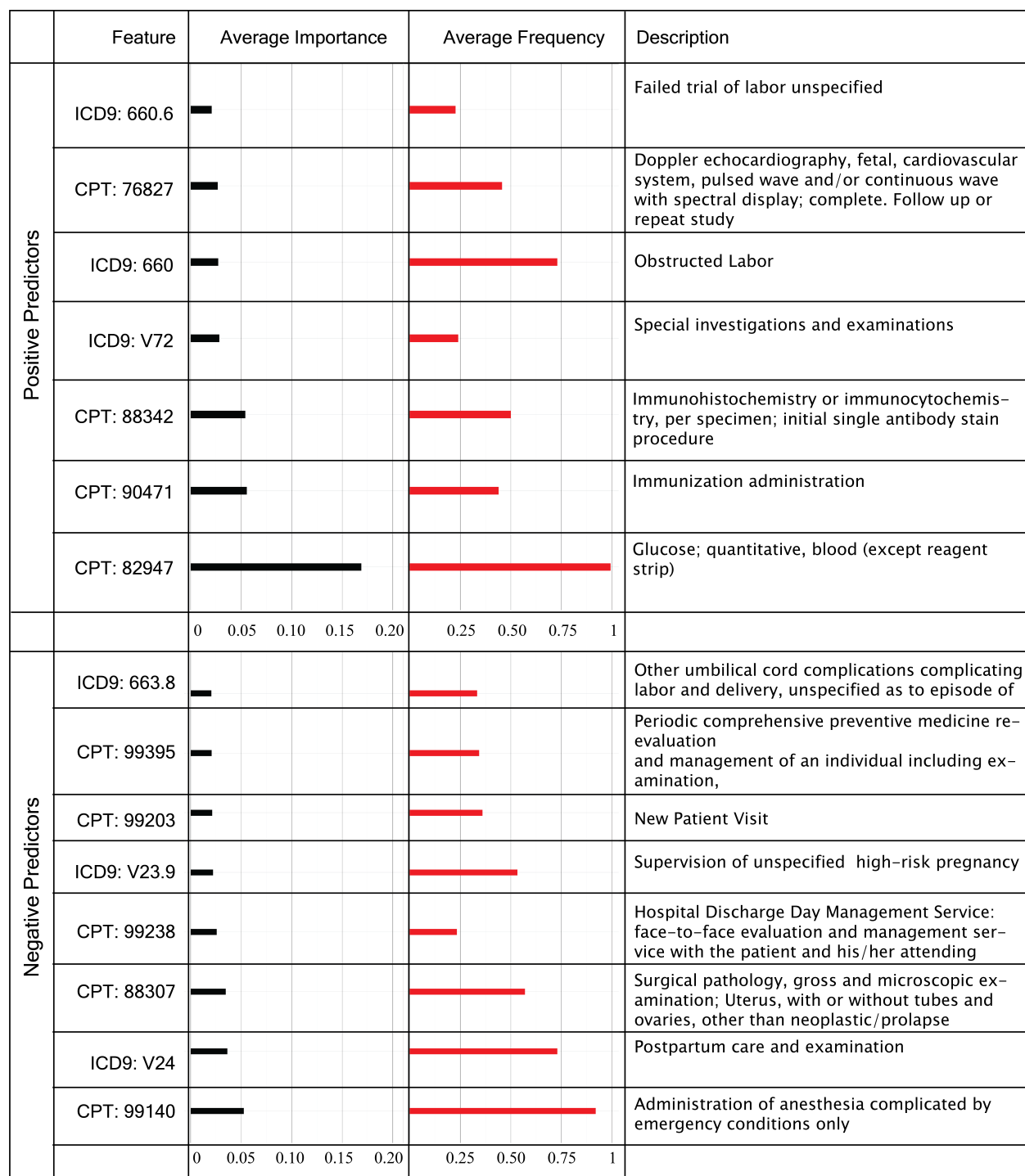
Our models also achieved an average precision of 0.8080 within a 0.8000 to 0.8159 95% confidence interval, an average sensitivity of 0.8079 within 0.7982 to 0.8177, and specificity of 0.8104 within 0.8014 to 0.8194.

### 3.2 Features Importance

On average, there were a small number of important features ( $\sim 50$ ) in each predictive model. Generally, over 90% features were eliminated in the models. To orient the audience, we only report on features that, on average, have an importance value greater than or equal to 0.05, and an average frequency value greater than or equal to 0.2. These two measures suggested 15 features, shown in Figure 3, for further investigation. Intuitively, features such as ICD-9 code 660.0 for obstructed labor by malposition of fetus at onset of labor and 660.6 for a failed trial of labor, make sense as positive predictors and are consistent with the finding that acute intrapartum events can lead to NE<sup>1</sup>. Note that the features that appear to represent observations after a birth, such as negative predictors ICD-9: V24 and CPT: 99238, are most likely from a previous delivery the mother had. These features that occur after a mother's previous deliveries are still considered prenatal features with respect to her subsequent deliveries.

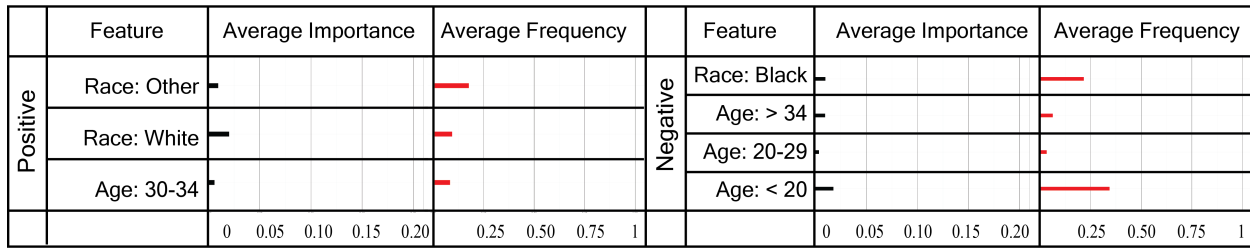
Age range and race category were included in the model to control for variability of age and race between case and controls. Importance and frequency of these potential confounding variables, shown in Figure 4, are relatively minimal, indicating that the distribution of age and race is similar between case and controls.

CPT code 82947 was a positive feature that exhibited a much higher average importance than any other feature and appeared in all 300 models. To verify that the cases were not dominated by this feature, we calculated the number of cases having that code and found only 16 cases that had this feature. Furthermore, we investigated if there were any



**Figure 3:** The 15 features with average frequency  $\geq 0.2$  and average importance  $\geq 0.05$ .

codes related with 82947 which appeared in the model. Given that CPT code billing procedure maintains that codes for related procedures are grouped by prefix, we identified codes with 829 as the first three digits and calculated the number of cases having each of them, as shown in Table 5. If these codes are correlated with each other, in the context



**Figure 4:** Average importance and frequency for Age Range and Race Category.

of the NE prediction objective, then all of our models will likely be dominated by the 829 procedure group. To further investigate, we measured the correlation between pairs of codes, and the correlation matrix between the CPT codes observed in the 829 group is visualized in Figure 5. From the figure, it can be seen that, surprisingly, little correlation exists between the codes aside from 82951 and 82952.

**Table 5:** The number of positive cases with an assigned code from CPT group 829.

CPT code	Cases with Code	Description
82947	16	Glucose; quantitative, blood (except reagent strip)
82950	69	Glucose; post glucose dose (includes glucose)
82951	14	Glucose; tolerance test, 3 specimens (includes glucose)
82952	15	Glucose; tolerance test, each additional beyond 3 specimens
82962	22	Blood glucose by glucose monitoring devices cleared by the FDA for home use
82977	4	Gamma Glutamyl Transferase

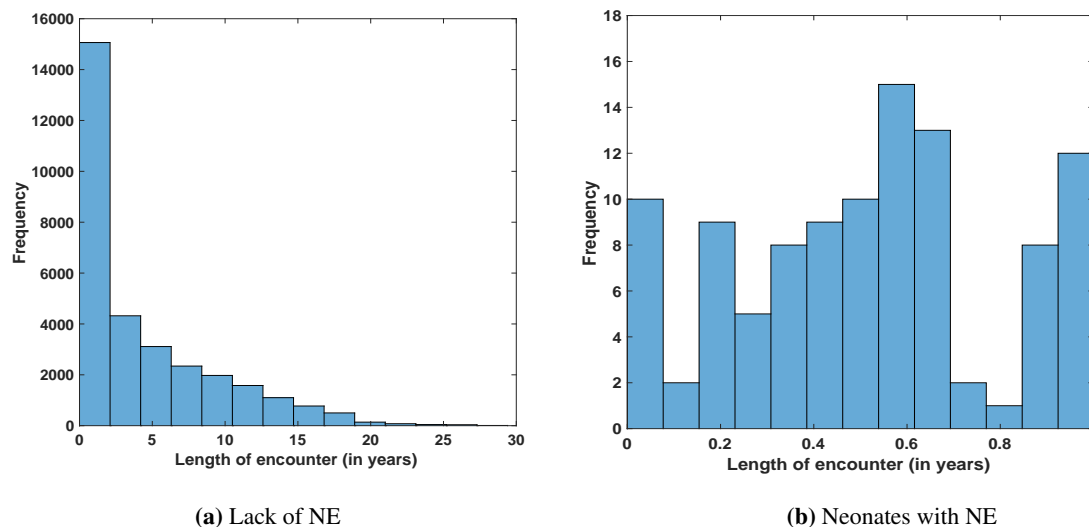
#### 4 Discussion

This work represents a prediction model of NE built around early maternal variables. Our prediction AUC of 87% on a relatively small number of features about an expecting mother is consistent with the hypothesis that cases of NE can be the result of distal risk factors, rather than acute intrapartum events. Since our model was trained with widely used medical billing codes, the same approach is easily replicable by other HCOs that are interested in modeling their maternal population.

Out of the selected features, a form of glucose test (CPT 82947) appeared as a highly important positive predictor. We offer two possible interpretations for this result. First, glucose tests would accompany a case of pregnancy induced diabetes, which could be a real risk factor for NE. Pregnancy induced diabetes is linked to fetal macrosomia<sup>22</sup>, a condition of high birth weight due to increased fetal glucose intake, and high birth weight is a well known risk factor for birth trauma and NE<sup>1,3,22</sup>. Secondly, glucose tests could be ordered for maternal patients new to VUMC. These new maternal patients may have been identified as predisposed to a high-risk delivery and redirected to VUMC, the premier medical institution in the area. The fact that all cases fall within a years length of encounter, discussed further below, supports this conclusion.

We acknowledge that this is a pilot study and has several limitations. First, the findings may be biased by the way in which high risk cases present to the VUMC. As the leading academic medical center in the area, VUMC receives patient cases that are often more complex or require more specialized care than other institutions to handle. As an artifact, many of the high-risk patients exhibited relatively short observational windows in the VUMC EMR. Figure 6a and 6b illustrates the length of observation history of mothers in the EMR before delivery windows for control and case patients, respectively. Mothers who had babies with NE had one year or less of medical history, which accentuates the concern that positive cases may have complications before receiving care at VUMC, introducing ascertainment bias into the data. Although we attempted to control for this confounding variable by collecting features within one year of the delivery in the controls, the models may, to some degree, be learning the distribution of diagnosis and procedures a patient receives after they are classified as high risk. Second, the number of NE cases in this study was relatively small and may not capture sufficient information of mothers with NE babies to perform prediction. The data observed was





**Figure 6:** Distribution of the length of observation (in years) for mothers whose baby (a) did not have NE and (b) those who did have NE. Only features that fell within one year prior to delivery were incorporated in our models.

taken from a single healthcare organization, and thus the patient population may be specific to VUMC. Our approach, however, can be applied with the data from any HCO that uses standardized billing terminologies.

Based on our findings and the limitations of the study, we outline the following improvements and extensions for the next iteration of this study. To begin, additional maternal and prenatal features can be investigated such as maternal lab values, in particular glucose values, and fetal ultrasound results. Medication data was previously included in the model, using the same binary feature construction scheme, but these features had low importance in the resulting model and no improvement in prediction was observed. Perhaps a non binary feature encoding may improve the signal received from medication prescriptions. Finally, we hope to understand the billing practices and clinical workflow behind the top features in order infer their potentially causal relationship with NE.

## 5 Conclusion

NE is a lethal condition that threatens millions of infants with premature death or lifelong morbidity. To date, the pathology of this complex affliction has eluded thorough understanding. Almost all existing NE studies are unable to perform a prenatal risk assessment of NE, making it challenging to adapt efficient interventions to prevent or treat NE in a timely manner. This paper introduces the first data-driven approach to predicting NE before an infant is born, and provides a mechanism to identify high-risk maternal factors for NE. The evaluation relied on a retrospective study of maternal EMR data at a large academic medical center. We considered a representation of the maternal condition centered around symptoms, diagnosis, and procedures, and trained classification models that consistently predict with high performance. However, this study is still preliminary in its scope and size. Further investigation and clinical interpretation of the reported features is required to understand the model, and replication of the study with other populations is needed to arrive at a better picture of the NE pathology. We believe this investigation opens up new and easily replicable approaches to understanding NE, and even may prelude a prediction system for recommending early NE interventions through the electronic health system.

## Acknowledgements

This research was supported, in part, by grants T15LM00740, R00LM011933, and R01LM010207 of the National Library of Medicine, National Institutes of Health.

## References

1. Academy College of Obstetricians and Gynecologists. Neonatal Encephalopathy and Neurologic Outcome, Second Edition. *Pediatrics*. 2014; 133(5): e1482-8.
2. Kurinczuk J, White-Koning M, Badawi N. Epidemiology of neonatal encephalopathy and hypoxicischaemic encephalopathy. *Early Hum Dev*. 2010; 86(6): 329-38.
3. Badawi N, Kurinczuk J, Keogh J, Alessandri L, O'Sullivan F, Burton P et al. Antepartum risk factors for newborn encephalopathy: the Western Australian case-control study. *BMJ*. 1998;317(7172): 1549-53.
4. Blume H, Loch C, Li C. Neonatal Encephalopathy and Socioeconomic Status. *Archives of Pediatrics & Adolescent Medicine*. 2007; 161(7): 663.
5. Bahubali G, Vishnu B, Ramachandra R, Nandakumar S, Adhisivam B, Rojo J et al. Antenatal and intrapartum risk factors for perinatal asphyxia: A case control study. *Curr Pediatr Res*. 2017; 17(2): 119-22.
6. Kaye D. Antenatal and intrapartum risk factors for birth asphyxia among emergency obstetric referrals in Mulago Hospital, Kampala, Uganda. *East Afr Med J*. 2004; 80(3): 140-3.
7. Milulescu A, Vladareanu S, Filipescu A, Mureanu N, Vladareanu R. Maternal risk factors for the neonatal hypoxic-ischemic encephalopathy. Preliminary results. *Romanian Society of Ultrasonography in Obstetrics and Gynecology*. 2017; 12: 135-8. References
8. Jones R, Heep A, Odd D. Biochemical and clinical predictors of hypoxicischemic encephalopathy after perinatal asphyxia. *The Journal of Maternal-Fetal & Neonatal Medicine*. 2017;;1-6.
9. Lundgren C, Brudin L, Wanby A, Blomberg M. Ante- and intrapartum risk factors for neonatal hypoxic ischemic encephalopathy. *The Journal of Maternal-Fetal & Neonatal Medicine*. 2017;;1-7.
10. Martinez-Biarge, M., Diez-Sebastian, J., Wusthoff, C., Mercuri, E. and Cowan, F. (2013). Antepartum and Intrapartum Factors Preceding Neonatal Hypoxic-Ischemic Encephalopathy. *PEDIATRICS*, 132(4), pp.e952-e959.
11. Landon M, Hauth J, Leveno K, Spong C, Leindecker S, Varner M, et al. Maternal and perinatal outcomes associated with a trial of labor after prior Cesarean delivery. *New Engl J Med*. 2004; 351(25): 2581-9.
12. Spong C, Landon M, Gilbert S, Rouse D, Leveno K, Varner M, et al. Risk of Uterine Rupture and Adverse Perinatal Outcome at Term After Cesarean Delivery. *Obstet Gynecol*. 2007;110(4):801-7.
13. Lilford R, Groot H, Moore P, Bingham P. The relative risks of caesarean section (intrapartum and elective) and vaginal delivery: a detailed analysis to exclude the effects of medical disorders and br *J Obstet Gynaecol*. 1990; 97(10): 883-92.
14. Yang X, Sun S. Comparison of maternal and fetal complications in elective and emergency cesarean section: a systematic review and meta-analysis. *Arch Gynecol Obstet*. 2017; 296(3): 503-12.
15. Cooke R. Chorioamnionitis, maternal fever, and neonatal encephalopathy. *Developmental Medicine & Child Neurology*. 2008;50(1):9-9.
16. Shalak L, Perlman J, Jackson G, Laptook A. Depression at Birth in Term Infants Exposed to Maternal Chorioamnionitis: Does Neonatal Fever Play a Role?. *Journal of Perinatology*. 2005;25(7):447-452.
17. Giuse DA. Supporting communication in an integrated patient record system. *AMIA Annu Symp Proc*. 2003: 1065.
18. Haibo He, Garcia E. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*. 2009; 21(9): 1263-84.
19. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005;67(2):301-320.
20. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Soft*. 2010; 33(1): 1-22.
21. Massey Jr FJ. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*. 1951; 46(253): 68-78.
22. KC, K., Shakya, S. and Zhang, H. (2015). Gestational Diabetes Mellitus and Macrosomia: A Literature Review. *Annals of Nutrition and Metabolism*, 66(2), pp.14-20.