Implementation of Multivariate Logistic Regression Model for Cerebral Palsy Identification using Prenatal, Perinatal Risk Factors

K. Muthureka¹, U. Srinivasulu Reddy²*, Mrs. B. Janet³

¹Research Scholar, Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India.
²*,³ Department of Computer Applications, National Institute of Technology, Tiruchirappalli, India.
*Corresponding Author Email ID: usreddy@nitt.edu

Abstract. Cerebral Palsy (CP), a static, neuro and motor disorder caused by brain injury in the time period of prenatal, perinatal and postnatal, is the major developmental disability affecting children’s function. Children with CP in children cannot be curable but quality of life can be improve with the help of treatment such as surgery and therapy. Early identification is important to the CP children for starting the treatment. There are numerous Machine Learning (ML) algorithms used in health care for prediction and classification. One of the ML algorithms called Logistic Regression which is used for binary classification using univariate and multivariate. This study, is of interest to enable early identification of CP using prenatal and perinatal risk factors with help of Multivariate Logistic Regression.

1. Introduction

CP is a neurodevelopmental and motor control disorder, caused by brain injury at the time of birth or at the early childhood [1]. Children with CP face a lot of problems [2-5] such as mobility, communicating with others, intellectual disability, retinal issues, epilepsy and most of the children suffer body pain. Nowadays, there are many ways to identify the CP in childhood through Birth History, Physical Examinations and Neuro Imaging. Diagnosing CP using Neuro Imaging is costly and developing countries provide limited facilities. Whereas Physical Examinations method need well-trained physicians. Finally, using Birth History is the easiest method of diagnose the CP.

Risk factors associated with CP can be grouped into three types [6,7] and these classification types are based on time period of child's birth, that is before birth called as prenatal, during birth called as perinatal and after birth called as postnatal. Figure 1 shows the classification of risk factor based on time period.
Perinatal asphyxia [8] is a major cause for CP. Birth asphyxia is a genuine global clinical issue and contributes enormously to morbidity and neonatal mortality [9]. Reasons for perinatal birth asphyxia might be maternal or fetal. The individuals who endure asphyxia during childbirth may have opportunity to create neurological entanglements including epilepsy, cerebral paralysis and formative postponement [10]. Another risk factor of perinatal period is birth weight. Most of preterm babies are born under weight. They may fall under two categories suggested by World Health Organization (WHO) as born with Low Birth Weight (LBW) and Very Low Birth Weight (VLBW). Babies weighing less than 2500 grams come under Low Birth Weight (LBW) category and weighing less than 1500 grams comes under Very Low Birth Weight (VLBW) category, some babies may extremely low weight which is less than 1000 grams as per WHO guidelines. During child birth, Apgar score calculated is prominent As it is directly related to birth weight and gestational age [11]. Van et al. [12] and his team had been identified in the literature work that gestational age, birth weight, gender, and medical diagnosis for the mother are the risk factor of CP.

Gestational Age (GA) is prenatal risk factor of CP. Children born in preterm are highly affected by CP disorder. The risk increases with decrease in gestational age [13-15]. The risk factor of cerebral palsy below 33 weeks gestation is 30 times greater than those who are born at term and is approximately equals to 70 in 1000 deliveries. Figure 2 presents classification of birth term based on gestation week under WHO advise. Early prediction of CP is highly essential inorder to provide better treatment to the make their living easier.

Machine Learning (ML) helps to make correct decisions based on observations and predictions. ML examines the areas of algorithms which makes high-end predictions on data accurately. ML automatically
improves through its experience. The training data is used to build the model and the testing data is used to validate the built model. Regression is one of the Supervised learning Algorithm of ML to find relationship between independent and dependent variables. The most popular Regression algorithm is based on regression value. The Logistic Regression (LR) model outcome variable is binary which makes it distinguishable from other models and hence, in the current work LR was adopted to early identification of CP using prenatal and perinatal risk factors. Mostly patients’ data are used to identify the influential attribute in predicting the given outcome [16].

2. Literature Review

Sukhov et al. [17] discussed about CP risk factors such as gestational age, age of mother, maternal race/ethnicity, twin, placenta abruption, mild to severe birth asphyxia, fetal distress, birth defects and birth trauma. This study found gestational age is risk factor of CP and also prematurity has the greatest impact upon the CP people’s future development.

Farin et al. [8] explained that the mother’s age is cause for the perinatal asphyxia (CP risk factor). In this study they used 159 cases and split into two group. Case group contains 53 participants and control group contains 106 participants. They found 6.6% of mothers under the age of 18 years at pregnancy, 90.8% of mothers in the 18–35 year age group, and 2.6% of mothers over the age of 35 years. Accuracy prediction of this study is 87.8%.

Linsell et al. [18] explained that children with very low birth weight or children born in very preterm (less than 32wks) have poor neuro development. This study suggests that very low birth weight and very preterm factors prognostic values of CP.

Luanda et al. [19] did research on 56 spastic CP children. In his study, he tried both Univariate and Multivariate Logistic Regression to identify the clinical and neurophysiologic characteristics associated with active transcranial direct current stimulation (tDCS) in CP children. This study took 7 weeks to complete and result was 30% increase on the 6 Minute Walk Test (MWT).

Deepta et al.[20] found that perinatal insult of children with CP causes Optic Nerve Head (ONH) morphology which can serve as an indicator for severe neurologic damage. In this study, authors have used 54 consecutive patients with CP and Multivariate Logistic Regression for research. The result of this study predicted 70% of all CP patients.

Tosun et al. [21] proposed to one of the machine learning algorithm, multivariate regression analysis to evaluate the relationship between BMD and possible risk factors. In this research, author’s team worked with various categories of children and the categories are 30 CP Children, 38 CP children with epilepsy, 54 children with Epilepsy only and 30 Healthy children. Team found Low BMD rate in children with Epilepsy only have 3.7%, only CP children have 50% and CP children with epilepsy have 39.5%.

Mann et al. [22] examined using multivariate regression analyses with 128 ambulatory CP children. In this study inspected the relationship between physical activity and walking performance to quality of life (QOL). Result of this study in the following manner physical QOL is 52.15, psychosocial QOL is 60.94 and total QOL is 56.55.

3. Materials and Methods

3.1 Data collection

Based on literature review we prepared questionnaire for data collection. We collected data from mothers of CP children and normal children. The questions were used to get information about child’s age, gender, mother’s medical history, gestational age, birth weight, asphyxia etc. We met parents of CP children at rehabilitation center and took a face to face interview. Table 1 shows the characteristics of participants.
### Table 1. Participant’s characteristics

|                      | Min -3 Years | Max -23 Years |
|----------------------|--------------|---------------|
| **Age**              | Female – 56% | Male - 44%    |
| **Gender**           |              |               |
| **Birth weight**     | Min – 1100g | Max – 3750g   |
| **Gestation age (weeks)** | Min -30    | Max – 40     |
| **Asphyxia**         | Yes – 36%   | No – 64%     |

### 3.2 Experimental setup

The current study to find the result using Multivariate analysis was conducted using LR. In the Multivariate analysis, risk factors were significantly associated with CP were taken as birth weight, gestational age and birth asphyxia. In order to build our proposed predictive model, we have exploited the capabilities of LR. The general formula of LR based on Sigmoid function can be expressed as follows.

\[
P = \frac{1}{1 + e^{-z}}
\]  

where \( P \) represents the probability of an event occurrence (dependent variable); \( z \) represents the independent variables linear combination function and \( z \) can also be expressed as follows.

\[
Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots \beta_n x_n
\]

where \( \beta_0 \) denotes the expected mean value of when \( x=0 \); \( x \) refers to the independent variables; \( \beta_n \) denotes to the regression coefficients of each independent variable and \( n \) represents the number of the independent variables.

By applying the linear combination function in equation (1) for the used tests as independent variables to determine an outcome (Cerebral Palsy Identification (CPI)); the formula of the predictive model that computes the probability of occurrence of CPI depending on the predictability of used tests can be written as below:

\[
P = \frac{1}{1 + e^{-[\beta_0 + \beta_1 T1 + \beta_2 T2 + \beta_3 T3]}}
\]

Where, \( P \) represents the probability of occurrence of an outcome (Cerebral Palsy Identification) based on the selected independent variables. \( \beta \) indicates to the regression coefficients of each included blood test. T1, T2, T3 refers to the included test parameters (“birth weight, gestational age and birth asphyxia”) respectively. We implement this work using python in jupyter notebook. The following section tells about implementing process.

### 3.3 Data analysis.

- First we have to preprocess the data. In preprocessing we encode the categorical values and we check the null value and visualize the location of missing values using seabornheatmap.
- Assign features where X contains birth weight, gestational age and birth asphyxia and Y contains affect and then split data into training and testing.
- After defining the model, we evaluate the performance using predict proba(), which returns the matrix of probabilities that the predicted result is equal to one or zero.

| 0.01918849 0.98081151 | 0.3353625 0.6646375 |
| 0.08370583 0.91629417 | 0.37049067 0.62950933 |
| 0.0507244 0.9492756 | 0.04963239 0.95036761 |
| 0.03174891 0.96825109 | 0.11093587 0.88906413 |
| 0.36928274 0.63071726 | 0.4675367 0.5324633 |
| 0.13281487 0.86718513 | 0.16790579 0.83209421 |
| 0.70203577 0.29796423 | 0.40454557 0.59545443 |
| 0.26975981 0.73024019 | 0.25609295 0.74390705 |
| 0.02649558 0.97350442 | 0.06930902 0.93069098 |
- Calculate the accuracy.

Accuracy of logistic regression classifier on test set: 0.83

- Receiver Operating Characteristic (ROC) curve is a useful tool for predicting the probability of a binary outcome. ROC curve used to plot the false positive rate in x-axis and true positive rate in y-axis based on number of different candidate threshold values between 0.0 and 1.0. Figure 3 indicates false positive rate and true positive rate.

![ROC Curve](image)

Figure 3. ROC Curve

### 4. Conclusion

An early identification is more important to the CP. We can easily find CP using causative risk factor. This study tried with minimum number of CP children’s birth history data which are collected directly from the mother. To the best of our knowledge, this is the first report of CP identification using birth weight; Gestational age and birth asphyxia with machine learning algorithm multivariate logistic regression and the accuracy is 83%.

### References

1. Zhang, B., Zhang, Y., 2008, Classification of cerebral palsy gait by Kernel Fisher Discriminant Analysis, *International Journal of Hybrid Intelligent Systems*, 5(4), pp 209–218.
2. Yesim Garip, Sumru Ozcel, Ozgul Bozkurt Tuncer, Gizem Kilinc, Feryal Seckin, Tansu Arasli, 2016, Fatigue in the mothers of children with cerebral palsy, *Disability and Rehabilitation*, pp 757-756.
3. Nimbalkar S, Raithatha S, Shah R, 2014, A qualitative study of psychosocial problems among parents of children with cerebral palsy attending two tertiary care hospitals in Western India, ISRN Family Medicine, 2014:769619.
4. Steven J. Korzeniewski, Jaime Slaughter, Madeleine Lenski, Peterson Haak, Nigel Paneth, September 2018, The complex aetiology of cerebral palsy, *Nature Review Neurology*, 14, pp 528-543.
5. Zeynep Alpay Savasan, Ali Yilmaz, Zafer Ugun, Buset Aydas, Ray O. Bahado-Singh Stewart F. Graham, 2 February 2019, Metabolomic Profiling of Cerebral Palsy Brain Tissue Reveals Novel Central Biomarkers and Biochemical Pathways Associated with the Disease: A Pilot Study, *Metabolites*.
6. Mahendra Rana, Jyoti Upadhayay, Amita Rana, Sumit Durgapal, Arvind Jantwal, 2017, A Systematic Review on Etiology, Epidemiology, and Treatment of Cerebral Palsy, *International Journal of Nutrition, Pharmacology, Neurological Diseases*, Wolters Kluwer Medknow, 7(4), pp 76-83.
7. Moshe Stavsky, Omer Mor, Salvatore Andrea Mastrolia, Shirley Greenbaum, Nandor Gabor Than, Ofri Erez, 2017, Cerebral Palsy Trends in epidemiology and Recent Development in Prenatal Mechanisms of Disease, Treatment, and Prevention, *Frontiers in Pediatrics*.
8. Soleimani F, Vamegh R, Biglarian A, 2013, Antenatal and Intrapartum Risk Factors for Cerebral Palsy in Term and Near-term Newborns, *Archives of Iranian Medicine*,
16(4), pp 213 – 216.

[9] Chayasak Pitsawong ,Prisana Panichkul, October 2011, Risk Factors Associated with Birth Asphyxia in Phramongkutklao Hospital , Thai Journal of Obstetrics and Gynaecology 19, pp 165-171.

[10] Haider BA, Bhutta ZA, 2006, Birth asphyxia in developing countries: current status and public health implications, Current Problems in Pediatric and Adolescent Health Care, 36(5), pp 178-188.

[11] Tove S. Rosen, David Bateman, 17 March 2010, Chapter 1 - The Effects of Gender in Neonatal Medicine , Principles of Gender-Specific Medicine (Second Edition) , pp 3-17.

[12] Van Lieshout P., Candundo, H. Martino. R., Shin,S, Barakat-Haddad, 2017, Onset factors in cerebral palsy: A systematic review, NeuroToxicology, 61, pp 47–53.

[13] Hafstrom M, Kallen K, Serenius F, 2018, Cerebral Palsy in Extremely Preterm Infants, Pediatrics, 141(1), e20171433.

[14] Himmelmann K, Uvebrant P, 2014, The panorama of cerebral palsy in Sweden. XI. Changing patterns in the birth year period 2003-2006, Acta Paediatrica, 103(6), pp 618–624.

[15] Oskoui M, Coutinho F, Dykeman J, Jetté N, Pringsheim T, 2013, An update on the prevalence of cerebral palsy: a systematic review and meta analysis, Developmental Medicine & Child Neurology, 55(6), pp 509–519.

[16] Evangelia Christodoulou, Jie Ma, Gary S. Collinsb, Ewout W. Steyerberg, Jan Y. Verbakel, Ben Van Calster, 2019, A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models, Journal of Clinical Epidemiology, 110, pp 12-22.

[17] Sukhov, A., Wu, Y., Xing, G., Smith, L. H., Gilbert, W. M, 2011, Risk factors associated with cerebral palsy in preterm infants, The Journal of Maternal-Fetal & Neonatal Medicine, 25, pp 53-57.

[18] Linsell, L., Malouf, R., Morris, J., Kurinczuk, J. J., Marlow, N, 2016, Prognostic factors For cerebral palsy and motor impairment in children born very preterm or very low birth weight : a systematic review, Developmental Medicine & Child Neurology, 58(6), pp 554-569.

[19] Luanda A. Collange Grecco, Claudia Santos Oliveira, Manuela Galli, Camila Cosmo, Natália de Almeida Carvalho Duarte, Nelci Zanon, Dylan J. Edwards Felipe Fregni, July 2016, Spared primary motor cortex and the presence of MEP in cerebral palsy dictate the responsiveness to tDCS during gait training, Frontiers in Human Neuroscience, 10, Article 361.

[20] Deepta Ghate, Vettaikorumakankav Vedanarayanan, Abdulbaset Kamour, James J. Corbett, Sachin Kedar, 2016, Optic nerve morphology as marker for disease severity in cerebral palsy of perinatal origin, Elsevier, pp 25-31.

[21] Ayse Tosun, Seda Erisen Karaca, Tolga Unuvar, Yakup Yurekli, Cigdem Yenisey, Imran Kurt Omurlu, 2016, Bone mineral density and vitamin D status in children with epilepsy, cerebral palsy, and cerebral palsy with epilepsy, Childs Nerves System, 33, pp 153–158.

[22] Kilby Mann, Elaine Tsao, Kristie F. Bjornson, 2016, Physical activity and walking Performance : influence on quality of life in ambulatory children with cerebral palsy (CP), Journal of Pediatric Rehabilitation Medicine: An Interdisciplinary Approach, 9, pp 279–286.