

## Cerebral Palsy Disorder Prediction Using Machine Learning Algorithms-A Review

Sarika Kshirsagar<sup>1</sup>, Sanjay Solanki<sup>2</sup>

<sup>1</sup>Assistant Professor, MCA DEPT, JSPM's Jayawantrao Sawant college of Engineering, Pune, Maharashtra, India

<sup>2</sup>Professor, MCA DEPT, JSPM's, Narhe Technical campus, Pune, Maharashtra, India

<sup>1</sup>kshirsagar.sarika@yahoo.com

<sup>2</sup>sanjay.solanki123@gmail.com

### Abstract

*Diagnosis is very important task in the field of medical science. Without this process patient cannot cured from any disease. now Machine learning techniques can help in medical science, ML predict the diseases with past history. after providing the details of patients it automatically predicts the diseases. hence it is very useful for doctors as well as patients. in this paper we provide review of prediction system using ML for cerebral palsy disorder which will help for researcher for predicting proposed system. we providing the review of recent papers in brief information and also comparing papers using some attributes.*

**Keywords:** cerebral palsy, Machine learning, review, prediction

### 1. Introduction

CP is the most common cause of motor disability in children and also affects a large adult population. The term cerebral palsy describes a group of movement disorders caused by an injury or disturbance in the early developing brain and specifically the areas involved with creating, coordinating and controlling movement and posture[1]. It is widely accepted that cerebral palsy does not result from a single cause but rather from a series of 'causal pathways' that can result in or accelerate injury to the developing brain. The cause of cerebral palsy in the majority of babies is still unknown. A recent systematic review of risk factors for cerebral palsy in children born at term in developing countries found 10 risk factors that were consistently reported as statistically significant predictors of cerebral palsy[2].

The 10 consistent risk factors identified include:

- Placental abnormalities
- Birth defects
- Low birth weight
- Meconium aspiration
- Instrumental/emergency caesarean delivery
- Birth asphyxia
- Neonatal seizures
- Respiratory distress syndrome
- Hypo glycaemia
- Neonatal infection.

Parent feel that this disorder is no cure, but if there will be proper diagnosis at earlier stage then children's can improve its situation. The machine learning can predict the disorder quickly with accurate result. lot of researchers worked on the prediction of CP disorder. they worked on various algorithms of ML. In this paper we present the literature review of machine learning algorithm for prediction and diagnosis of CP.

## 2. Overview of Machine Learning Techniques

Machine learning (ML) is a domain of artificial intelligence that involves constructing algorithms that can learn from experience. The way that ML algorithms work is that they detect hidden patterns in the input dataset and build models. Then, they can make accurate predictions for new datasets that are entirely new for the algorithms. This way the machine became more intelligent through learning; so, it can identify patterns that are very hard or impossible for humans to detect by themselves. ML algorithms and techniques can operate with large datasets and make decisions and predictions[3]. Machine learning techniques are of 3 types

- 1]Supervised learning
- 2]Unsupervised learning
- 3]reinforcement learning

In supervised learning, the target is to infer a function or mapping from training data that is labeled[4]. In unsupervised learning, we lack supervisors or training data. In other words, all what we have is unlabeled data. The idea is to find a hidden structure in this data. There can be a number of reasons for the data not having a label, the reinforcement learning method aims at using observations gathered from the interaction with the environment to take actions that would maximize the reward or minimize the risk[4].

Supervised machine learning is used in prediction as compare to unsupervised and reinforcement learning. Classification is main type of supervised learning. It consist of following algorithms:

### 2.1. Decision tree (DT):

This algorithm has a tree-like structure or flowchart-like structure. It consists of branches, leaves, nodes and a root node. The internal nodes contain the attributes while the branches represent the result of each test on each node. DT is widely used for classification purposes because it does not need much knowledge in the field or setting the parameters for it to work[3].

### 2.2. Support vector machine (SVM)

SVM is a concept in statistics and computer science for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the input, making the SVM a non-probabilistic binary linear classifier[5].

### 2.3. Artificial neural network

A neural network is a model that is designed by the way human nervous systems such as brain, that process the information. Neural networks, with their remarkable ability to derive meaning

from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. Many neural network models, even biological neural networks assume many simplifications over actual biological neural networks [5].

#### **2.4. K-nearest neighbor (KNN)**

In KNN classification, the output is a class membership. An object is classified by a plurality vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours (k is a positive integer, typically small). If k=1, then the object is simply assigned to the class of that single nearest neighbour[6]

#### **2.5. Logistic regression**

It is estimating the parameters of a logistic model; it is a form of binomial regression. Mathematically, a binary logistic model has a dependent variable with two possible values, such as pass/fail, win/lose, alive/dead or healthy/sick; these are represented by an indicator variable, where the two values are labeled "0" and "1".[6]

#### **2.6. Random forest**

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees habit of overfitting to their training set[6].

### **3. Review of Different Prediction System for Cerebral Palsy Disorder Using Machine Learning**

In this paper we provide the literature review of prediction algorithm for cerebral palsy disorder. we collected these papers according to ML algorithms which is used in prediction the model.

#### **3.1 Decision tree**

Stoen et al.[7]used the decision tree algorithm for early CP prediction based on infant's video recording.they collected the data from Norwegian university hospital,405 infants recording used for prediction with at least age 18-24 months. accuracy of the predicted model is 95.3%. limitations of this study is it focused only CP outcomes of participants. not included perspective of parents of high-risk children. dataset included only infants.

Trost et al.[8] developed and tested decision tree model to classify physical activity intensity from accelerometer output and gross motor function classification system in ambulatory youth with CP and compare classification accuracy of new decision tree models to that previously published cut points for youth with CP. 57 ambulatory youth information taken as a dataset. Models were trained and cross validated using R. Accuracy of predicted model is 84%. Small amount of dataset and dataset includes only youth is the limitation of the study.

Borgne et al.[9]developed data mining tool for prediction of CP using decision tree. Data is collected from two hospitals in Brussel and gilletteinsti in Minnesota. Gait data are recorded from motion analysis system (e.g., Vicon and BTS). R and python apache server tools used for

data processing and data testing. Accuracy of the model is 90%. predicted model considered only gait index.

### 3.2 Logistic regression

Bertoncelli et al.[10]used logistic regression algorithm for prediction of model.130 children's data collected and their age in between 12-18 years. Data is between June 2005-June 2015.for data processing and analyzing using R software is used. accuracy of the model is 90%. author predict the model for small amount of data.

Hermanson et al.[11] developed the predicted model using logistic regression .data collected from Sweden .145 children between 3-5 years age taken for dataset.87% is the accuracy of predicted model. They studied only GMFCS III to V level included hip displacement.

Anish et al.[12] studied student enrolment in education and its determinants with CP using binary logistic regression analysis. They collected data from medical college thiruvantapuram, Kerala between 3-12 years age children. chi- square test is used for analysis of data. Seizure disorder might have contributed to the failure to the schooling which is not included in analysis. Predicted model accuracy is 95%.

### 3.3 Hybrid algorithms

Bahado-singh et al.[13] used 5 algorithms SVM,LDA,PAM,GLM and logistic regression for prediction of CP in newborns.23 leucocyte DNA of newborn used for testing. all data collected from Michigan department of health and human services .For developing the model R software is used. SVM gave the 95% accuracy. They used ML technique to predict CP in newborns but there is no longer validation exist.

Ahmadi et al.[14]created prediction regression model using RF,SVM and BDT.22 children of 6-20 years of age is considered for dataset.89% accuracy of this hybrid approach. there are some limitations that it is not evaluated in free living environment. less number of data and there are limited set of features like brisk walking, comfortable walking etc.

Ihlen et al.[15] used partial least square regression algorithm for prediction of CP in infants.377 infants of 9-15 weeks of age tested for prediction. model is called CIMA (computer-based infants movement assessment).

## 4.Comparison of Different Algorithm for Cerebral Palsy Disorder Prediction

Here we compare the review papers in tabular form. we consider 7 attributes as follows. we compare the accuracy as well as its limitations.

- 1.Year: It shows the year of publication of research paper or thesis or book
- 2.Author: It shows name of author and refence number
- 3.Techniques: It shows which technique is used like classification techniques, its type .it can be single algorithm or hybrid algorithms
- 4.Dataset: It is required as input for any algorithm.

5. Tools: it shows software or programming language for processing the dataset and training, testing the dataset. predict the model using these tools.

6. Accuracy: It shows that how much predicted model is accurate and its result in the form of accuracy

7. Limitations: It represent the limitation of the study.

**Table 1. Comparison of different algorithm for cerebral palsy disorder prediction**

Year	Author	Technique/ algorithms	Dataset	Tool	Accuracy	Limitations
2020	Bertoncelli et al.[10]	Logistic regression	130 children , 12-18 age	R	90%	Study on very small of data.
2019	Bahado-singh et al.[13]	SVM, LDA, PAM, GLM, LR	23 cases of CP	R	95%	The study is only for newborn
2019	Stoen et al.[7]	Decision tree	405 infants, 18-24 month age	SPSS	95.3%	It focuses on CP outcome of participant
2019	Ihlen et al.[15]	Partial least square Regression, LDA	377 infants, 9-15 weeks	Not mentioned	87%	Study for infants only.
2018	Ahmadi et al.[14]	RF, SVM, BDT	22 children, 6-20 years age	F test	89%	Limited no of features included like brisk walking, very less dataset
2016	Trost et al.[8]	Decision tree	57 ambulatory youth	R	84%	Dataset included only youth, very less amount of data.
2015	Hermanson et al.[11]	Logistic regression	145 children's , 3-5 years age	Not mentioned	87%	Study for GMFCS III to V level. feature of hip displacement is considered for study.
2013	Borgne et al.[9]	Decision tree	Two hospital data in Minnesota	R, Python, apache server	90%	It works only on gait index.
2013	Anish et al.[12]	Logistic regression	Medical college, kerala, 3-	Chi square test	95%	Seizure disorders management

			12 years age			contributed to failure which is not included in analysis.
--	--	--	--------------	--	--	---

## 5. Conclusion

This paper overviews the literature of cerebral palsy disorder prediction using machine learning techniques. Most of the researchers used decision tree and logistic regression machine learning algorithms. Hybrid approach is also used for prediction. Tool used for processing the algorithm is R software. From above study we conclude that Stoen et al. developed the predicted model have high accuracy than other models. i.e. decision tree algorithms gave the highest accuracy. Hybrid approach is very useful in prediction the diseases. SVM and logistic regression algorithms also gave the best accuracy. In above studies we find that dataset is not available on UCI repository or any data websites. All authors collected their data from hospitals or rehabilitation centers. So preparing the quality dataset is very important task for getting best accuracy of algorithm.

## References

- [1] F. Diagnosis and T. O. Understanding, "THE CEREBRAL PALSY."
- [2] *Management Of Cerebral Palsy In Children : A Guide For Allied Health Professionals MANAGEMENT OF CEREBRAL PALSY IN CHILDREN - A GUIDE FOR ALLIED HEALTH PROFESSIONALS.*
- [3] M. Aljanabi, M. Qutqut, and M. Hijawi, "Machine Learning Classification Techniques for Heart Disease Prediction: A Review," *Int. J. Eng. Technol.*, vol. 7, no. October, pp. 5373–5379, 2018, doi: 10.14419/ijet.v7i4.28646.
- [4] M. Mohammed, M. B. Khan, and E. B. M. Bashie, *Machine learning: Algorithms and applications*, no. December. 2016.
- [5] G. B.M, S. C.P, and S. T, "Breast Cancer Diagnosis Using Machine Learning Algorithms - A Survey," *Int. J. Distrib. Parallel Syst.*, vol. 4, no. 3, pp. 105–112, 2013, doi: 10.5121/ijdps.2013.4309.
- [6] K. Nirmaladevi, S. Shanthi, T. Agila, R. T. Dharani, and P. Dhivyapriya, "Analysis and Prediction of Diabetes using Machine Learning," *Test Eng. Manag.*, vol. 83, pp. 14533–14538, 2020.
- [7] R. Støen et al., "The Predictive Accuracy of the General Movement Assessment for Cerebral Palsy: A Prospective, Observational Study of High-Risk Infants in a Clinical Follow-Up Setting," *J. Clin. Med.*, vol. 8, no. 11, p. 1790, 2019, doi: 10.3390/jcm8111790.
- [8] S. G. Trost, M. Fragala-Pinkham, N. Lennon, and M. E. O'Neil, "Decision trees for detection of activity intensity in youth with cerebral palsy," *Med. Sci. Sports Exerc.*, vol. 48, no. 5, pp. 958–966, 2016, doi: 10.1249/MSS.0000000000000842.
- [9] Y. Le Borgne, G. Bontempi, and B. Bonnechere, "Data mining toolbox for gait analysis in children with cerebral palsy," *Isb*, no. August 2019, pp. 1–3, 2013, [Online]. Available: <http://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/151884/Details>.
- [10] C. M. Bertonecelli, P. Altamura, E. R. Vieira, S. S. Iyengar, F. Solla, and D. Bertonecelli, "PredictMed: A logistic regression-based model to predict health conditions in cerebral palsy," *Health Informatics J.*, 2020, doi: 10.1177/1460458219898568.
- [11] M. Hermanson, G. Hägglund, J. Riad, E. Rodby-Bousquet, and P. Wagner, "Prediction of hip displacement in children with cerebral palsy: Development of the cpup hip score," *Bone Jt. J.*, vol. 97-B, no. 10, pp. 1441–1444, 2015, doi: 10.1302/0301-620X.97B10.35978.
- [12] T. Anish, R. Ramachandran, P. Sivaram, S. Mohandas, A. Sasidharan, and P. Sreelakshmi, "Elementary school enrolment and its determinants among children with cerebral palsy in Thiruvananthapuram district, Kerala, India," *J. Neurosci. Rural Pract.*, vol. 4, no. 5 SUPPL, pp. 40–44, 2013, doi: 10.4103/0976-3147.116446.
- [13] R. O. Bahado-Singh, S. Vishweswaraiyah, B. Aydas, N. K. Mishra, C. Guda, and U. Radhakrishna, "Deep learning/artificial intelligence and blood-based DNA epigenomic prediction of cerebral palsy," *Int. J. Mol. Sci.*, vol. 20, no. 9, 2019, doi: 10.3390/ijms20092075.
- [14] M. Ahmadi, M. O'Neil, M. Fragala-Pinkham, N. Lennon, and S. Trost, "Machine learning algorithms for activity recognition in ambulant children and adolescents with cerebral palsy 11 Medical and Health

- Sciences 1103 Clinical Sciences,” *J. Neuroeng. Rehabil.*, vol. 15, no. 1, pp. 1–9, 2018, doi:  
10.1186/s12984-018-0456-x.
- [15] E. A. F. Ihlen *et al.*, “Machine Learning of Infant Spontaneous Movements for the Early Prediction of Cerebral Palsy: A Multi-Site Cohort Study,” *J. Clin. Med.*, vol. 9, no. 1, p. 5, 2019, doi:  
10.3390/jcm9010005.