**Predicting child literacy-numeracy, physical, learning and social-emotional development in Bangladesh Using Machine Learning Approach**

1. **Introduction:**

Early childhood development is crucial for lifelong health and well-being, offering both significant growth opportunities and vulnerability to harm (1). Key aspects of child development, such as self-regulation, early relationships, knowledge acquisition, and skill development, are shaped by neurobiology, caregiver interactions, and the caregiving environment, influencing a child's learning readiness and future health outcomes (2–4). The first 1,000 days of life, from conception to age two, are crucial for physical, social, emotional, and cognitive development. Experiences during this period profoundly influence a child's growth and future development (5).

Developmental delay in children occurs when a child significantly lags behind peers in reaching milestones related to motor skills, speech, social skills, daily living activities, and cognition. It is a descriptive term rather than a diagnosis and usually applies to children under 5 but can extend into later childhood or even adulthood (6,7). It can be isolated (affecting a single domain), multiple (affecting two or more domains), or global (impacting most developmental areas). Developmental delays can result from poor birth outcomes, inadequate stimulation, malnutrition, chronic health issues, organic problems, psychological and familial factors, or other environmental influences (8). While developmental delay may not be permanent, it can provide a basis for identifying children who may experience a disability Early detection and intervention are crucial to prevent long-term disability (9). This underscores the importance of early identification to initiate timely interventions with family involvement, aiming to prevent delays, foster emerging skills, and create a more stimulating and protective environment (8).

The exact prevalence of developmental delay is unclear, but the WHO estimates that 10% of the population in each country has some form of disability (10). Globally, 52.9 million children under five have developmental disabilities, mostly in low- and middle-income countries (95%), while 15% of U.S. and 2.17% of children under five in England have developmental issues (11) (9). In 2016, a study updated the estimate of the number of children at risk of poor development. It concluded that approximately 43% of children under 5 years of age living in low- and middle-income countries (LMIC), more than 250million children under age 5 years in low/middle-income countries (LMICs) are at risk of not attaining their full development potential (12). Between 2010 and 2016, 25.3% of children in 63 low and middle-income countries (LMICs) had a developmental deficit, with 10.1% in Europe and Central Asia, 32.6% in South Asia, 17.0% in East Asia and Pacific, and 41.4% in West and Central Africa experiencing developmental delays (13). In Bangladesh, the Multiple Indicator Cluster Surveys (MICS) 2012 found 30% of children were not developmentally on track, while the MICS 2019 reported 25.26% were not meeting milestones according to the Early Childhood Development Index (ECDI) (14).

Children’s development is influenced by a wide range of biological and environmental factors, some of which protect and enhance their development while others compromise their developmental outcomes (8). Children in poverty face higher developmental delays than those from wealthier backgrounds due to increased exposure to risks and long-term effects on brain development from poverty and trauma (15,16). Inadequate nutrition and poor maternal nutritional status during pregnancy can lead to intrauterine growth restriction, adversely affecting brain development and highlighting the crucial role of maternal nutrition in fetal survival, growth, and development (17). Early childhood programs are crucial for supporting young children's mental and physical development, with participants showing significantly better developmental progress compared to their peers (15). Adequate supervision, stimulation, and access to books and toys positively impact early childhood development (ECD) status. Nurturing care enhances health and development, while neuroscientific data indicates it can counteract the effects of poor socioeconomic conditions on brain development. Conversely, child punishment, physical abuse, family instability, risky neighborhoods, and poverty can lead to poor coping skills, emotional regulation issues, and lower social functioning in children (18,19).

Child developmental outcomes are complex and arise from intricate interactions between biological, environmental, and social factors, often poorly understood. Traditional statistical methods, which rely on unrealistic assumptions, may struggle to model this complexity and hinder research into risk prediction tools(20). Statistical models, which depend on predefined relationships and can be affected by missing values and outliers, may be less effective in complex scenarios. However, machine learning, leveraging advances in data from birth cohorts, health records, imaging, and other sources, offers new potential for modeling these interactions and identifying optimal predictive patterns for early interventions (21).

Traditional statistical methods focus on inferring relationships based on predefined assumptions, while machine learning (ML) offers improved prediction and reveals complex patterns that traditional methods may miss (22). The purpose of machine learning (ML) is to make repeatable predictions by learning patterns from data without relying on predefined assumptions or rules (23,24). Machine learning (ML) does not infer causal relationships but learns patterns from training data to make predictions on new data. It can process vast amounts of diverse data, including epidemiological, imaging, environmental, and genomic data, efficiently handle missing values and outliers, and manage feature redundancy and selection. ML models are non-parametric, excelling at identifying significant variables for prediction (25) (26). A major advantage of machine learning is its ability to manage numerous highly interactive predictors and non-linear relationships, which can enhance predictive accuracy as data quantity and diversity increase (27).

Machine learning does not always guarantee better predictive performance than traditional methods. Effectiveness depends on the outcome, the number and type of features, and feature interactions. It's important to trial and report traditional methods like logistic and linear regression before considering more complex ML models (28). This work introduces a machine learning approach using multilabel classifiers to analyze childhood developmental data and compares cost-effective prediction models for developmental delays using a large MICS survey sample. We evaluated seven multilabel classifiers and identified the most relevant features for prediction.

References:

1. Shonkoff JP, Garner AS, Siegel BS, Dobbins MI, Earls MF, McGuinn L, et al. The Lifelong Effects of Early Childhood Adversity and Toxic Stress. Pediatrics [Internet]. 2012 Jan 1 [cited 2024 Aug 11];129(1):e232–46. Available from: /pediatrics/article/129/1/e232/31628/The-Lifelong-Effects-of-Early-Childhood-Adversity

2. Wesarg-Menzel C, Ebbes R, Hensums M, Wagemaker E, Zaharieva MS, Staaks JPC, et al. Development and socialization of self-regulation from infancy to adolescence: A meta-review differentiating between self-regulatory abilities, goals, and motivation. Developmental Review. 2023 Sep 1;69:101090.

3. Gee DG, Cohodes EM. Caregiving Influences on Development: A Sensitive Period for Biological Embedding of Predictability and Safety Cues. Curr Dir Psychol Sci [Internet]. 2021 Oct 10 [cited 2024 Aug 11];30(5):376. Available from: /pmc/articles/PMC8528226/

4. Brennan LM, Shelleby EC, Shaw DS, Gardner F, Dishion TJ, Wilson M. Indirect effects of the Family Check-Up on school-age academic achievement through improvements in parenting in early childhood. J Educ Psychol. 2013 Aug;105(3):762–73.

5. Likhar A, Baghel P, Patil M. Early Childhood Development and Social Determinants. Cureus [Internet]. 2022 Sep 23 [cited 2024 Aug 11];14(9). Available from: /pmc/articles/PMC9596089/

6. Bellman M, Byrne O, Sege R. Developmental assessment of children. BMJ [Internet]. 2013 Jan 15 [cited 2024 Aug 11];346(7891). Available from: https://www.bmj.com/content/346/bmj.e8687

7. Tonelli M, Parkin P, Leduc D, Brauer P, Pottie K, Jaramillo Garcia A, et al. Recommendations on screening for developmental delay. CMAJ [Internet]. 2016 May 17 [cited 2024 Aug 11];188(8):579–87. Available from: https://www.cmaj.ca/content/188/8/579

8. Early Childhood Development and Disability: A discussion paper.

9. Allen L, Ornstein M. Developmental Delay. Practical Pediatric and Adolescent Gynecology [Internet]. 2023 Jul 17 [cited 2024 Aug 11];158–62. Available from: https://www.ncbi.nlm.nih.gov/books/NBK562231/

10. Dornelas LDF, Duarte NMDC, Magalhães LDC. [Neuropsychomotor developmental delay: conceptual map, term definitions, uses and limitations]. Rev Paul Pediatr [Internet]. 2015 Mar 1 [cited 2024 Aug 11];33(1):88–103. Available from: https://pubmed.ncbi.nlm.nih.gov/25662016/

11. Salomone E, Pacione L, Shire S, Brown FL, Reichow B, Servili C. Development of the WHO Caregiver Skills Training Program for Developmental Disorders or Delays. Front Psychiatry [Internet]. 2019 Nov 11 [cited 2024 Aug 11];10. Available from: /pmc/articles/PMC6859468/

12. Lu C, Black MM, Richter LM. Risk of poor development in young children in low-income and middle-income countries: an estimation and analysis at the global, regional, and country level. Lancet Glob Health [Internet]. 2016 Dec 1 [cited 2024 Aug 11];4(12):e916–22. Available from: https://pubmed.ncbi.nlm.nih.gov/27717632/

13. Gil JDC, Ewerling F, Ferreira LZ, Barros AJD. Early childhood suspected developmental delay in 63 low- and middle-income countries: Large within- and between-country inequalities documented using national health surveys. J Glob Health [Internet]. 2020 Jun 1 [cited 2024 Aug 11];10(1). Available from: /pmc/articles/PMC7295453/

14. Progotir Pathey. Bangladesh multiple indicator cluster survey 2019 Key findings. 2019;

15. Hasan MN, Babu MR, Chowdhury MAB, Rahman MM, Hasan N, Kabir R, et al. Early childhood developmental status and its associated factors in Bangladesh: a comparison of two consecutive nationally representative surveys. BMC Public Health [Internet]. 2023 Dec 1 [cited 2023 Aug 30];23(1):1–13. Available from: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-023-15617-8

16. Luby JL. Poverty’s Most Insidious Damage: The Developing Brain. JAMA Pediatr [Internet]. 2015 Sep 1 [cited 2024 Aug 11];169(9):810–1. Available from: https://jamanetwork.com/journals/jamapediatrics/fullarticle/2381540

17. Herring CM, Bazer FW, Johnson GA, Wu G. Impacts of maternal dietary protein intake on fetal survival, growth, and development. Exp Biol Med [Internet]. 2018 Mar 1 [cited 2024 Aug 11];243(6):525. Available from: /pmc/articles/PMC5882021/

18. Noble KG, Houston SM, Brito NH, Bartsch H, Kan E, Kuperman JM, et al. Family income, parental education and brain structure in children and adolescents. Nature Neuroscience 2015 18:5 [Internet]. 2015 Mar 30 [cited 2024 Aug 11];18(5):773–8. Available from: https://www.nature.com/articles/nn.3983

19. Mares ML, Pan Z. Effects of Sesame Street: A meta-analysis of children’s learning in 15 countries. J Appl Dev Psychol. 2013 May 1;34(3):140–51.

20. Barker LE, Shaw KM. Best (but oft-forgotten) practices: checking assumptions concerning regression residuals. Am J Clin Nutr [Internet]. 2015 Sep 1 [cited 2024 Aug 12];102(3):533–9. Available from: https://pubmed.ncbi.nlm.nih.gov/26201816/

21. Dwyer D, Koutsouleris N. Annual Research Review: Translational machine learning for child and adolescent psychiatry. Journal of Child Psychology and Psychiatry [Internet]. 2022 Apr 1 [cited 2024 Aug 11];63(4):421–43. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/jcpp.13545

22. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health. 2018;8(2).

23. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods [Internet]. 2018 Apr 3 [cited 2024 Aug 12];15(4):233–4. Available from: https://pubmed.ncbi.nlm.nih.gov/30100822/

24. Azzolina D, Baldi I, Barbati G, Berchialla P, Bottigliengo D, Bucci A, et al. Machine learning in clinical and epidemiological research: isn’t it time for biostatisticians to work on it? Epidemiol Biostat Public Health [Internet]. 2019 [cited 2024 Aug 12];16(4). Available from: https://riviste.unimi.it/index.php/ebph/article/view/17117

25. Mafi S, AbdelRazig Y, Doczy R. Machine Learning Methods to Analyze Injury Severity of Drivers from Different Age and Gender Groups. https://doi.org/101177/0361198118794292 [Internet]. 2018 Aug 29 [cited 2024 Aug 12];2672(38):171–83. Available from: https://journals.sagepub.com/doi/10.1177/0361198118794292

26. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med [Internet]. 2016 Sep 29 [cited 2024 Aug 12];375(13):1216–9. Available from: https://pubmed.ncbi.nlm.nih.gov/27682033/

27. Bowe AK, Lightbody G, Staines A, Murray DM. Big data, machine learning, and population health: predicting cognitive outcomes in childhood. Pediatr Res [Internet]. 2023 Jan 1 [cited 2024 Aug 12];93(2):300. Available from: /pmc/articles/PMC7614199/

28. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol [Internet]. 2019 Jun 1 [cited 2024 Aug 12];110:12–22. Available from: https://pubmed.ncbi.nlm.nih.gov/30763612/