

## ORIGINAL ARTICLE

# Ranking Predictors of Child Dietary Diversity in Nepal Using a Decision Tree

Dipesh Baral<sup>1</sup>, Dindu Lama<sup>2</sup>, Archana Dangal<sup>3</sup>, Dinesh Baral<sup>4</sup>, Jiban Kharel<sup>4</sup>, Abhishek Thapa<sup>4</sup>, Pragati Dahal<sup>5</sup>, Swastika Baral<sup>4</sup>

## ABSTRACT

**Introduction:** Minimum dietary diversity among children remains low in Nepal, contributing to persistent malnutrition. This study aimed to rank and describe associations of dietary, socioeconomic, and health-related predictors with minimum dietary diversity among children aged 6–59 months in Madhyapur Thimi Municipality, Nepal.

**Methods:** A secondary analysis was conducted using survey data from 375 children. Baseline associations were examined using chi-squared tests, and a chi-squared-based decision tree classification method was fitted to identify key split variables and interaction structures. A model-based sensitivity analysis estimated changes in predicted dietary adequacy when individual food groups were switched from not consumed to consumed. Bootstrap resampling quantified internal variability in these estimates.

**Results:** Overall, 73.6% of children were predicted to achieve minimum dietary diversity. The decision tree identified consumption of other fruits and vegetables (Group G) as the primary split, followed by eggs (Group E), legumes, and nuts (Group B), and vitamin A-rich foods (Group F). Sensitivity analysis suggested that increasing consumption of Group G or Group E produced the largest expected improvements in predicted dietary adequacy, and bootstrap results indicated that this ranking was relatively stable.

**Conclusion:** This model-based analysis indicated that Group G and Group E foods are strong positive predictors of adequate dietary diversity and offer the largest expected gains among diet-deficient children. Hence, access to Group G and Group E foods could be prioritized as dietary interventions to improve child nutrition.

**Keywords:** Bootstrap, Children, Decision tree, Nutrition, Sensitivity analysis

<sup>1</sup>Department of Mathematics and Statistics, Washington State University, Washington, USA

<sup>2</sup>Department of Business and Economics, Wittenberg University, Ohio, USA

<sup>3</sup>Registered Nurse, Paradise Creek Health and Rehabilitation of Cascadia, Idaho, USA

<sup>4</sup>Independent Researcher, USA

<sup>5</sup>Department of Agricultural and Applied Economics, Virginia Polytechnic Institute and State University, Virginia, USA

**Corresponding author:**

Dipesh Baral, Department of Mathematics and Statistics, Washington State University, Pullman, WA.

Email: [dipesh.baral@wsu.edu](mailto:dipesh.baral@wsu.edu)

## INTRODUCTION

Several studies have examined child nutrition and dietary diversity in Nepal and have highlighted persistent malnutrition, but the key determinants emphasized differ across settings.<sup>1-4</sup> Minimum dietary diversity for children is defined as consumption of at least four of seven standard food groups (excluding breastfeeding), categorized by the World Health Organization (WHO) in the previous day, and is a widely used indicator of child feeding adequacy.<sup>5</sup> Nationally, fewer than half of Nepali children aged 6–59 months meet this standard, and low dietary diversity is linked to poor growth and development.<sup>6-8</sup> Prior work in Madhyapur Thimi identified socio-demographic and behavioral correlates, but it did not rank predictors or quantify the contribution of specific food groups.<sup>3</sup> To support local prioritization, this study applied a chi-squared-based decision tree approach to rank dietary and non-dietary predictors of dietary adequacy and to identify high-leverage food groups for improving minimum dietary diversity.

## METHODS

This study was a secondary analysis of a cross-sectional survey conducted in 2023 among 385 children aged 6–59 months in wards 2 (Jatigaal) and 3 (Kaushaltar) of Madhyapur Thimi Municipality, Nepal. Details on the original sampling, instruments, data collection procedures, and ethical review have been published previously.<sup>3</sup> Ten children with missing outcome data were excluded, resulting in an analytic sample of 375. Because all children consumed grains, roots, and tubers (Group A), that food-group indicator was not used as a predictor due to zero variation.

Ethical approval for the parent survey was granted by the Institutional Review Committee of Nobel College, Nepal, and local administrative permissions were obtained before fieldwork.<sup>3</sup> Mothers/guardians were informed of the study purpose, that participation was voluntary, and that they could withdraw at any time without penalty. Consent was documented in writing after verbal agreement (signature or tick mark when needed). Privacy was maintained during interviews, and data was managed to protect confidentiality and

anonymity. The present study was a secondary analysis of a de-identified dataset, conducted with authorization for secondary use from the original investigators.

The outcome was minimum dietary diversity adequacy, defined as consumption of at least four of the seven standard food groups in the previous day and coded as adequate (Yes) versus inadequate (No).<sup>5</sup> Predictors included dietary food-group indicators—legumes and nuts (Group B), dairy products (Group C), flesh foods (Group D), eggs (Group E), vitamin A-rich fruits and vegetables (Group F), and other fruits and vegetables (Group G)—measured as consumed versus not consumed. Additional predictors covered socio-demographic characteristics (child sex and age; maternal age, education, and ethnicity; household income; monthly food expenditure; occupation; family type; household size), health-related factors (maternal awareness of communicable diseases; current child illness such as common cold, diarrhea, pneumonia, skin infection, or other), and feeding/behavioral factors (daily feeding frequency; maternal handwashing before cooking; maternal awareness of ultra-processed foods; and recent ultra-processed food consumption, including items such as momo, samosa, chowmein, noodles, potato chips, and ice cream, along with a composite indicator of any ultra-processed foods).

A complete-case analysis was performed (N = 375). Descriptive statistics for this sample have been reported previously.<sup>3</sup> Associations between each predictor and dietary adequacy were assessed using chi-squared tests with a significance threshold of  $p < 0.05$ . A decision tree was then fitted using the Chi-squared automatic interaction detection (CHAID) procedure, with dietary adequacy as the dependent variable and dietary, socio-demographic, health, and feeding predictors as candidate split variables. Splits were selected using chi-squared tests with Bonferroni-adjusted p-values ( $\alpha = 0.05$ ), with minimum node sizes of 10 for further splitting and 5 for terminal nodes. A model-based sensitivity analysis was conducted to estimate the expected change in predicted adequacy when individual food-group indicators were switched from not consumed to consumed among eligible children; this quantity was summarized using the

expected flip potential (EFP) metric, as defined in the Supplementary Materials. Internal variability and stability of the food-group rankings were evaluated using nonparametric bootstrap resampling (B = 500), refitting the tree in each resample and summarizing the mean expected flip potential, percentile intervals, and the proportion of resamples in which each food group ranked highest. All analyses were conducted in R version 4.3.1. Additional mathematical details are provided in the Supplementary Materials.

### RESULTS

Of the 375 children, 73.6% were predicted to achieve minimum dietary diversity among children (MDD-C). Bivariate chi-square tests identified 12 significant predictors ( $p < 0.05$ ), with dietary variables dominating (Table 1). The top-ranking predictors were Food Group G ( $\chi^2 = 108.94$ ,  $p = 0.00050$ ), Food Group B ( $\chi^2 = 59.60$ ,  $p = 0.00050$ ), Food Group F ( $\chi^2 = 48.67$ ,  $p = 0.00050$ ), and Food Group E ( $\chi^2 = 39.84$ ,  $p = 0.00050$ ). The leading non-dietary predictors were maternal awareness of communicable diseases ( $\chi^2 = 23.01$ ,  $p = 0.00050$ ), educational level of the mother ( $\chi^2 = 18.60$ ,  $p = 0.00350$ ), and monthly expenditure on food ( $\chi^2 = 14.46$ ,  $p = 0.00100$ ).

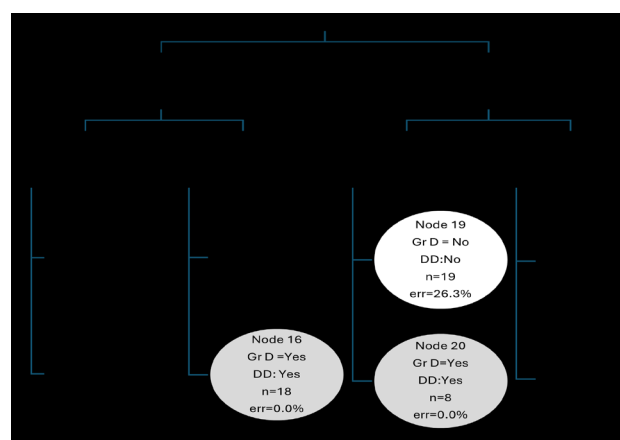
**Table 1: Ranked significant predictors of MDD-C (chi-square test,  $p < 0.05$ ,  $N=375$ ) and their CHAID tree splitting positions**

Rank	Variable	$\chi^2$ statistic	$\chi^2$ p-value	Level in Tree
1	Group G	108.94	0.00050	0 (root)
2	Group B	59.60	0.00050	1-2
3	Group F	48.67	0.00050	2-3
4	Group E	39.84	0.00050	1
5	Maternal Awareness of Communicable Disease	23.01	0.00050	NA
6	Mother's educational level	18.60	0.00350	NA
7	Group D	14.55	0.00050	2-3
8	Monthly Expenditure on Food	14.46	0.00100	NA

9	Group C	13.15	0.00100	3
10	Maternal Awareness of UPF	10.35	0.00100	NA
11	Mother's Washing Hand Behavior	5.48	0.03398	NA
12	Status of Common Cold in Children	5.33	0.02699	NA

Note: "First level in tree" is the depth at which a predictor first appears as a splitting variable (root = 0). NA indicates that the Variables did not appear as splits.

The CHAID model comprised 31 nodes (15 inner; 16 terminal leaves) across 5 levels, with dietary factors dominating. A portion of the CHAID tree is visualized in Figure 1, while the whole tree is available in the Supplementary material. The root split was on Group G. Key dietary branches included Group E, Group F, Group B, and Group D. Non-dietary splits (e.g., family income, UPF consumption such as samosa) appeared only at deeper levels.



**Figure 1: CHAID Decision Tree for MDD-C Predictors.**

Note: DD = MDD-C outcome (yes/no); n = number of observations in the node; err = misclassification error within the node; Gr = food group indicator (e.g., Gr G= Group G). Circular nodes are terminal nodes, and rectangular nodes are intermediate nodes. Shaded nodes indicate nodes predicting DD: Yes.

**Table 2: Model-based expected flip potential (EFP) and bootstrap uncertainty**

Group	Mean EFP (%)	EFP 95% CI	Top-rate
Group G	62.6	47.5–76.3	0.55
Group E	55.7	26.2–81.1	0.34
Group F	52.7	28.0–70.4	0.10
Group B	23.7	11.0–36.6	0.00
Group D	18.5	2.2–38.3	0.01
Group C	9.6	0–18.3	0.00

*Note. EFP = expected flip potential. Percentages are expressed relative to the effective number of children predicted to have inadequate dietary diversity. Top-rate is the proportion of bootstrap samples in which a food group achieved the highest EFP percentage among deficient children.*

From table 2, it can be seen that adding Group G, Group E or Group F foods to the diets of currently deficient non-consumers would be expected to move roughly half or more of the deficient group into adequate dietary diversity (Group G: 62.6%, 95% bootstrap CI 47.5–76.3%; Group E: 55.7%, 26.2–81.1%; Group F: 52.7%, 28.0–70.4%). Other food groups had markedly smaller effects, with a mean below 25%. Across bootstrap resamples, Group G, Group E, and Group F were almost always the leading levers (top-rates 0.55, 0.34, and 0.10, respectively), whereas other groups almost never ranked first.

**DISCUSSION**

The predicted 73.6% prevalence of MDD-C adequacy observed in Madhyapur Thimi is nearly identical to prior findings of 73.5%, indicating consistency with earlier estimates from the same setting.<sup>3</sup> As expected, dietary variables overwhelmingly dominated both bivariate chi-squared tests and the CHAID model, reflecting the definitional role of food group consumption in dietary diversity.

CHAID is a supervised machine learning classification algorithm based on recursive partitioning with chi-squared tests.<sup>9</sup> It is particularly useful for identifying nonlinear relationships and hierarchical interactions among categorical predictors, making it well-suited for variables such as food group consumption and

socio-demographic characteristics.<sup>9</sup> Previous studies have applied CHAID to investigate dietary behavior and public health outcomes.<sup>10-16</sup> However, to our knowledge, no prior study has used this method to rank predictors of diet diversity in Nepal.

Unlike the binary MDD-C threshold ( $\geq 4$  groups), which weighs all food groups equally, the CHAID model revealed a clear hierarchy: Group G emerged as the most significant determinant, followed by Group E, Group F, and Group B. Group C and Group D appeared only at deeper nodes and showed only small non-zero gains in the sensitivity analysis, suggesting their secondary roles. Tree-based clusters, sensitivity analysis, and bootstrap statistics all illustrate these dynamics. For example, the largest inadequate subgroup (Node 4 in the tree) lacked Group G, Group E, and Group F foods. Similarly, adding Group G, Group E, or Group F to the diets of all currently predicted diet-deficient non-consumers was associated with the largest expected gains, each expected to move more than half of the deficient subgroup into MDD-C adequacy (Table 2) in the sensitivity analysis. Furthermore, bootstrap statistics indicated that one of these three groups was almost always the leading lever across bootstrap resamples. This structure underscores that interventions should not merely encourage “any additional group” but prioritize high-impact categories—particularly Group G, Group E, and Group F.

Regarding non-dietary predictors, we first note that several such variables that were significant in the bivariate screening, such as maternal awareness of communicable disease, maternal education, and maternal handwashing practices, were never selected as splits, suggesting that their influence was largely captured by the food-group. Second, we note that even when they appear, they only appear in deeper splits, again underscoring their secondary role relative to dietary variables. For example, household income appeared only at Nodes 29 and 30, after the tree had already split on Groups G, B, and C. Similarly, consumption of UPFs entered the tree only in specific branches (e.g., Samosa at Nodes 14, 15, 28, and 31), where they mainly worsened predictions in subgroups already lacking key food groups. Taken together, these patterns illustrate how CHAID can map hierarchical

interactions, with non-dietary factors such as income and UPF intake conditioning, rather than replacing the dominant influence of core food groups.

The findings advance the original study in four key ways. First, the analysis identified the most influential food groups and key non-dietary factors, whereas the earlier study excluded dietary variables. Second, the predictors were hierarchically organized in the CHAID decision tree rather than only tested in isolation, allowing their relative importance to be compared within the same model. Third, the marginal improvement in dietary diversity from each food group was quantified through the expected flip potential (EFP) analysis. Fourth, bootstrap resampling characterized the internal variability and ranking stability of these model-based results. Together, these contributions not only extend prior findings but also illustrate a practical way to apply supervised machine-learning methods to public health nutrition research.

The findings suggest two practical pathways for intervention. First, dietary strategies should prioritize accessibility to Group G, Group E, and Group F foods over other groups, as they consistently emerged as the most influential groups across the tree and flip analyses. Approaches such as subsidized fruit markets, egg distribution, and school feeding programs centered on these foods could form the cornerstone of community nutrition efforts. Second, maternal awareness of communicable disease and the pattern of ultra-processed food consumption indirectly influenced child nutrition outcomes through their position in deeper splits. Targeted nutrition education workshops to strengthen awareness and promote healthier behaviors would therefore complement dietary interventions and enhance their effectiveness.

This analysis has some limitations. This is a model-based simulation analysis of a CHAID tree, and control of confounding variables and assumptions of causal inference have not been explicitly examined. Additionally, the underlying dietary diversity score is ordinal, but we dichotomized it at the MDD-C threshold, so information was lost when collapsing it to a binary Yes/No outcome. Such ordinal data is ubiquitous; however, it remains unclear whether CHAID fully respects this ordinality.<sup>17-19</sup> Finally, dietary

diversity was assessed via recall, which may be subject to reporting bias.

## CONCLUSION

Dietary diversity components—particularly Group G, Group E, and Group F—emerged as the strongest predictors of MDD-C in Madhyapur Thimi. Non-dietary factors, including maternal awareness of communicable diseases, awareness of ultra-processed foods, and socioeconomic conditions, also contributed but played a secondary role. The hierarchical structure of the CHAID model provides a practical framework for designing sequential interventions. Prioritizing access to key food groups (Group G, Group E, and Group F) combined with awareness campaigns and broader socioeconomic support may help improve child dietary diversity in similar settings and could contribute to reducing child malnutrition.

## ACKNOWLEDGMENT

Data used in this secondary analysis were collected as part of the previously published parent study.<sup>3</sup> Grammarly and ChatGPT were used for language editing (grammar, punctuation, and sentence clarity) only. The authors are responsible for the study design, analyses, interpretation, and final manuscript. Secondary analysis was conducted in accordance with the parent study's data-use provisions.

## CONFLICT OF INTEREST

The authors declare no conflict of interest. Data and analysis codes will be made available upon reasonable request.

## REFERENCE

1. Acharya I, Acharya A, Pant S, Pradhan PM. Nutritional Status of Under-five Children and its Associated Factors in Dhulikhel Municipality, Kavrepalanchok, Nepal: Nutritional Status in Children of Dhulikhel Municipality. *Nepal Journal of Public Health*. 2025 Jun 9;2(1):21-6. DOI: [https://doi.org/10.70280/njph\(2025\)v2i1.29](https://doi.org/10.70280/njph(2025)v2i1.29)
2. Singh AK, Acharya SP, Pandey A. Nutritional Status in 6-59 months Children and its Association with Water, Sanitation and Hygiene Practice in Panchapuri Municipality, Surkhet, Karnali Province,

- Nepal. Nepal Journal of Public Health. 2025 Jun 9;2(1):41-9. DOI: [https://doi.org/10.70280/njph\(2025\)v2i1.24](https://doi.org/10.70280/njph(2025)v2i1.24)
3. Thapa A, Baral D, Dahal P, Kharel J, Dangal A, Baral S, et al. Dietary diversity and its associated factors among children aged 6-59 months in Madhyapur Thimi Municipality, Nepal. Nepal Journal of Public Health. 2025 Jun 9;2(1):50-6. DOI: [https://doi.org/10.70280/njph\(2025\)v2i1.30](https://doi.org/10.70280/njph(2025)v2i1.30)
  4. Shrestha S, Khanal V. Dietary practices among young children in Nepal's Karnali Mountains: a community-based study on meal frequency, diversity, and fruit or vegetable intake. BMC Public Health. 2025;25(1):2114. DOI: <https://doi.org/10.1186/s12889-025-23312-z>
  5. World Health Organization. Indicators for assessing infant and young child feeding practices: conclusions of a consensus meeting held 6-8 November 2007 in Washington, D.C., USA. Geneva: WHO; 2008.
  6. Ministry of Health, Nepal; New ERA; ICF. Nepal demographic and health survey 2016. Kathmandu: Ministry of Health, Nepal; 2017.
  7. Vijay J, Patel KK. Malnutrition among under-five children in Nepal: a focus on socioeconomic status and maternal BMI. Clin Epidemiol Glob Health. 2024;27:101571. DOI: <https://doi.org/10.1016/j.cegh.2024.101571>
  8. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. The lancet. 2013;382(9890):427-51. DOI: [10.1016/S0140-6736\(13\)60937-X](https://doi.org/10.1016/S0140-6736(13)60937-X)
  9. Kass GV. An exploratory technique for investigating large quantities of categorical data. J R Stat Soc Ser C Appl Stat. 1980;29(2):119-27. DOI: <https://doi.org/10.2307/2986296>
  10. Bôto JM, Marreiros A, Diogo P, Pinto E, Mateus MP. Health behaviours as predictors of the Mediterranean diet adherence: a decision tree approach. Public Health Nutr. 2022;25(7):1864-76. DOI: [10.1017/S1368980021003293](https://doi.org/10.1017/S1368980021003293)
  11. Chavis C, Bhuyan IA. Data-driven food desert metric to understand access to grocery stores using chi-square automatic interaction detector decision tree analysis. Transp Res Rec. 2022;2676(5):571-89. DOI: <https://doi.org/10.1177/03611981221089308>
  12. Bozkir AS, Sezer EA. Predicting food demand in food courts by decision tree approaches. Procedia Comput Sci. 2011;3:759-63. DOI: <https://doi.org/10.1016/j.procs.2010.12.125>
  13. Ribas Seguí D, Forcadell MJ, Vila-Córcoles A, de Diego-Cabanes C, Ochoa-Gondar O, Martin Lujan F, et al. Classification rule for ten year MACE risk in primary care Tarragona older adults with type 2 diabetes: a CHAID decision-tree analysis. BMC Prim Care. 2025;26(1):129. DOI: <https://doi.org/10.1186/s12875-025-02826-w>
  14. Hanslian E, Schiele JK, Jeitler M, Michalsen A, Wischnowsky M, Storz MA, et al. Attitudes toward healthy nutrition in Germany—results from a representative cross-sectional online survey. Front Nutr. 2025;11:1480980. DOI: <https://doi.org/10.3389/fnut.2024.1480980>
  15. Hong S, Kim HW, Walton B, Kaboi M. The intersectionality of factors predicting co-occurring disorders: a decision tree model. Int J Ment Health Addict. 2024 Jul 22. DOI: <https://doi.org/10.1007/s11469-024-01358-1>
  16. Goldstein A, Kapelner A, Bleich J, Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J Comput Graph Stat. 2015;24(1):44-65. DOI: <https://doi.org/10.1080/10618600.2014.907095>
  17. Baral D, Lama D, Baral D. Predicting Low Mathematics Grades from Math Anxiety Using Weighted Ordinal Logistic Regression. American Journal of STEM Education. In Press.
  18. Bista R, Poudel P, Baral D. Examining the influence of demographic variables on mathematics test anxiety. American Journal of STEM Education. 2025; 3:105-19. DOI: <https://doi.org/10.32674/8kr0sr83>
  19. Baral D, Dasgupta N, McLeman L, Bista R, Jones J, Biswakarma M. Exploring the predictive relationship between students' math anxiety and their mathematical performance. American Journal of STEM Education. 2025; 3:1-15. <https://ojed.org/STEM/article/view/7422>