

From Basic Indicators To Direct Health Status: A Multi-Model Machine Learning Approach For Malnutrition Detection In Children

Dr. J. Karunanithi¹, E. Sundaravalli²

¹Associate Professor & Research Head Dept. of computer science, TSA AST College Perur, Coimbatore-10, cschod.jk@gmail.com

²Research scholar Dept. of computer science TSA AST College Perur, Coimbatore-10, esundaravalli23@gmail.com

Abstract

Malnutrition in children is one of the serious global health concerns, especially after COVID-19. The lower and middle-income countries are the most affected and highly susceptible to malnutrition problems. The inadequate access to nutritious food, healthcare facilities, clean water, intervention programs and early childhood care and education are the main causes of the childhood malnutrition crisis. Malnutrition in children under the age of five affects their physical, cognitive and emotional behaviour and development. Furthermore, it reduces their productivity in the long term, which leads to health-related problems, reduced academic performance and economic hardship. Even though various initiatives and intervention services are introduced and implemented globally, malnutrition continues to affect an overwhelming number of children under the age of five. The immediate necessities for innovative and data-driven strategies are required to improve the early malnutrition detection, prevention and treatment processes across the world. In this era, machine learning is one of the revolutionary technologies in health care systems to build data-driven decisions and targeted interventions in diagnostic and treatment processes. Machine learning algorithms facilitate to analysis of diverse datasets such as anthropometry features, clinical and developmental factors, socio-economic and environmental indicators to detect the malnutrition health status of the children. The different types of machine learning models were developed, such as linear and non-linear models, ensemble methods and boosting methods using Logical Regression, Decision Tree, Random Forest, Gradient Boosting and XGBoost classification algorithms to classify the early malnutrition health statuses of the children. These ML models were enhanced by the Synthetic Minority Oversampling Technique (SMOTE) to reduce class imbalances and a hyperparameter tuning process to reduce overfitting issues. These techniques were effective in this prediction system to enhance its prediction accuracy. The Gradient Boosting machine learning model outperforms other machine learning models with the highest accuracy.

INTRODUCTION

Malnutrition is the state of imbalance between the actual nutritional needs of the body and the intake or observed nutritional levels. Malnutrition can be divided into undernutrition and overnutrition. Calorie deficiency, insufficient protein and micronutrient conditions referred to as undernutrition and excess nutrient consumption are referred to as overnutrition. Stunting, wasting, and underweight are the key indicators of malnutrition. These indicators are essential to assess the nutritional health status of children under the age of five. The global health monitoring programs utilise these indicators to monitor the child's nutritional health status.

The most crucial phases of children are 0-2 years and 1-5 years. The 0-2 years (the first 1000 days) phase is the most critical for a child's physical and mental growth. The formation of a strong and healthy immune system, 80% of the brain development and irreversible nutrition impacts are the key developments in this stage. This stage is the foundation for children's life to acquire their long-term health, intelligence and productivity. It is important to encourage breastfeeding and weaning practices, growth monitoring and prevention of stunting and wasting conditions during this phase. The 1-5 years phase (toddlers phase) is important for physical growth, cognitive skills, learning, and social and emotional development of the toddlers. The infection risks, underweight, and micronutrient deficiencies are the major challenges of this stage. These challenges can be prevented by focusing on a balanced diet with micronutrients, a safe environment, early education, vaccinations, deworming process, sanitation and hygiene lifestyles.

The malnutrition impacts on physical, cognitive and immune system growth in children under five.

- Malnutrition impacts on physical growth: poor nutrition in children leads to poor physical development, such as stunting, underweight and wasting. These physical growth delays cause delayed motor development, weak growth of muscles and bones, low energy levels, and less active and sensitive to fatigue. In extreme cases, malnutrition leads to impairment in the development of significant organs, overall health and productivity.
- Malnutrition impacts on cognitive development: The malnutrition in toddlers affects their brain development with structural variations in the regions (memory, attention and learning) and reduced brain volumes. The nutrients such as iron, iodine, zinc and fatty acids help in neurotransmitter production and brain signalling. These micronutrient deficiencies lead to delayed language skills, poor academic results and lower IQs. The children with malnutrition face multiple challenges with concentration, critical thinking, problem-solving and emotional balance.
- Malnutrition impacts on immune system functions: malnutrition reduces the body's defence mechanisms and results in children being more vulnerable to infections – diarrhoea, pneumonia and measles. These infections affect the production and function of immune system cells and compromise the vaccine's efficacy. This makes them more severe, long-lasting and frequent illness and further weakens their nutritional levels. It is the compromised immunity that leads to the mortality risks. The children with stunted growth and compromised immune systems, or overnutrition, cause the expansion of obesity and associated health issues. There is a mutual relationship between the immune system and malnutrition. [25]

The malnourished children under five years are susceptible to both short-term and long-term health-related problems. Cognitive delays and learning difficulties can affect their academic progress. In their adulthood, they may face lower productivity, lower earning potential and fewer opportunities for their financial advancement. Chronic diseases like hypertension, diabetes, and cardiovascular disorders are the major problems due to early malnutrition and metabolic imbalances. It is important to address the multi-level causes of malnutrition among individuals and society.

- Immediate causes: The immediate or direct causes of malnutrition are inadequate nutritional food, infectious diseases like diarrhoea, maternal health, child's feeding practices, birth weight and accessibility to the health services.
- Underlying causes: The underlying causes of malnutrition are due to the household and community level aspects that affect the nutritional status of the child. The poor sanitation and living style make the unhygienic household environment, inconsistent access to nutritious food leads to food insecurity, and lower-income facility enables the inability to access basic needs.
- Basic causes: The basic causes of malnutrition are related to political, environmental and socio-economic conditions of the society. Maternal mental health, education, gender equality, poverty level, food scarcity, agriculture, climate, government policies and schemes related to health and food security are the basic causes of malnutrition in children in society. [26]

Due to the subtle or non-specific symptoms of wasting, stunting, underweight or swelling, the delayed recognition of malnutrition in children has been a big problem in their later stages of life. The symptoms cannot be visible until it has progressed significantly. It causes early nutrition deficiencies, which impact long-term growth. Several factors contributing to the delayed diagnosis of malnutrition are socio-cultural norms normalising the symptoms with the genetic derivations, poor appetite with the common habits and inconsistent malnutrition screening processes. The table below [26] explains the direct, indirect and cross-cutting intervention strategies of nutrition enhancements in children.

Direct Intervention Strategies	Indirect Intervention Strategies	Cross-Cutting Intervention Strategies
Breast feeding support	Household food security	Strengthening the health care systems
Supplementary foods	Women's empowerment	Data systems management
Nutritional programs	Water sanitation	Community mobilization
Micronutrient enrichment	Access to education	Performance tracking
Anaemia treatment	Eradication of poverty	Ensuring transparency

TABLE I. Intervention Strategies for children's Nutrition Enhancements

The improvisation of early malnutrition detection can be achieved by the regular growth monitoring, awareness programs, health education and utilising the evolving technologies like machine learning or

deep learning. Several existing methods are available for the detection of malnutrition: using anthropometry measurements and clinical examinations. The anthropometry measurements include weight-for-age (WFA), height-for-age Z-scores (HAZ), and Mid-Upper Arm Circumference (MUAC). The weight-for-age (WFA) factor compares the child's weight with the average weight. The average weight is measured as the average weight of children of the same age. This metric is commonly used to identify the underweight status. It may neither differentiate between acute and chronic malnutrition nor depict the exact nutritional status. Height-for-Age Z-scores (HAZ) metric assesses the stunting condition (a sign of chronic malnutrition). The below -2 of Z-score exhibits the shorter height compared with the average height of children of the same age. The standard tools and accurate age are essential to identify the HAZ. But these are not always available in the field. Mid-Upper Arm Circumference (MUAC) is a basic and immediate measurement to identify Severe Acute Malnutrition (SAM). It is measured by the circumference of a child's upper arm. It focuses on muscle mass and fat and is effective in most of the critical cases. Still misses detecting the chronic malnutrition. The physical signs of malnutrition are oedema, hair discolouration, and dermatosis. These symptoms are used to detect protein-energy malnutrition, such as kwashiorkor or marasmus. The identification of these health issues can be done by experienced healthcare workers. Though various methods are found and applied in real-life scenarios, some limitations are associated with these methods. The trained professionals are required to utilise the measurement approaches, tools and result interpretations. These methods are not always available in remote regions and rural areas. Manual assessments are always subjective, and they lead to inaccuracy risks. This one-dimensional approach (weight, height, or arm circumference) may not consider the risks associated with the dietary intake, clinical signs, socio-economic status and environmental aspects of factors. These limitations can be overcome by integrating the multi-dimensional data with the machine learning models. Early and accurate malnutrition detection can be achieved by implementing machine learning models in the malnutrition detection process.

Machine learning techniques are powerful and game-changing tools in various fields, especially in healthcare systems. Advanced prediction, classification, and decision support mechanisms are possible in the current scenario through machine learning approaches. The machine learning algorithms detect the hidden patterns in the given dataset, the relationship among the attributes and risk features associated with the target feature. It enables the data-driven approach and the capability of working with large volumes of data.[27] In the field of healthcare systems, machine learning models can analyse a complex and wide range of patient information, such as clinical measurements, demographic details, and lifestyle factors, to predict diseases and suggest personalised treatment procedures. Several machine learning techniques were implemented in the early detection of diseases, and treatment plans are helpful in the patient's fast recovery from the illness. Depending on the input features and algorithms, different accuracy levels are possible. [28]

This predictive mechanism makes machine learning significant in the field of malnutrition prediction, especially where early detection and multiple factors integration are required. The machine learning models enabling faster and smarter decision-making support to address the limitations of the current methodologies. The real-time decision-making approach assists healthcare workers in early identification of high-risk cases and prioritising individuals who require immediate intervention. It provides better outcomes for susceptible populations. The advantages of implementing machine learning models for malnutrition detection are:

- Early detection and intervention: The machine learning systems can recognise the mild patterns and predict the malnutrition status even before the visible clinical symptoms. It is possible to make timely interventions (nutritional or medical) according to the outcomes of the ML-systems.
- Integrating a diverse set of data types: The ML-based models can analyse a broad range of input features - anthropometric feature sets, clinical signs and socio-economic factors, simultaneously. It can capture holistic advantages of complex and different feature sets.
- Learning complex relationships: The machine learning models uncover the complex relationships among the feature sets better than human experts or conventional statistical methods.
- Automation in screening: Once the machine learning model is trained with the sample data with limited human interventions, the large-scale monitoring in community health programs can be done efficiently.

- Minimal dependency on human experts: The automation of pattern detection and decision-making support of machine learning models leads to reduced hiring of healthcare workers and human errors or bias. This is particularly useful for rural areas or remote regions.
- Continuous learning and improvement: The ML systems are able to learn continuously with the newly updated data over time. The accuracy and adaptation to unseen datasets will be improved.
- Support for resource allocation: it is easy to assess the higher risk populations or regions to allocate the resources, like healthcare workers, food and other resources.

The primary objective of this study is to implement an enhanced machine learning model that accurately classifies the malnutrition health status of children between 0 and 5 years. This enhanced machine learning model applies the multi-source and multi-dimensional dataset, such as anthropometric measurements, clinical and developmental indicators and socio-economic and environmental indicators. The analysis of these multi-dimensional data makes the machine learning model an early warning system to predict and prioritise the high-risk cases. This data-driven strategy suggests proactive interventions to the policy makers, which prevent the long-term consequences of both physical and mental consequences related to the early malnutrition of the children.

LITERATURE REVIEW

Children under five years old are the most susceptible to health and nutrition issues. Especially the first 2 years of the child's life are the crucial part of their entire life. Good nutrition is the core that is associated with their growth and development. The improper nutrition intake can affect their immune system and overall growth. [1] The United Nations Sustainable Development Goals prioritise “zero hunger” as Goal 2 to eradicate malnutrition in children under the age of five. [2]

The developing countries deal with the undernutrition problems very commonly. The undernutrition problems make the children with a weakened immune system, which allows a high risk of infections and metabolic diseases. The escalation of this existing state has happened during COVID-19 and resulted in huge breakdowns in health and socio-economic conditions around the world. Only 54% of the children were moderately or seriously malnourished before COVID-19. The situation of COVID-19 caused excess health problems related to malnutrition. Some of the statistics are

- 2.6 million children were affected by stunting,
- 9.3 million were affected by wasting,
- 2.1 million maternal anaemia cases were detected,
- 168,000 child mortality,
- On the other hand, Productivity loss occurred for USD 29.7 billion.

These incidents occurred due to a shortage of food, harvest disruptions, stress in the food supply chain and cost, which led to improper nutrition intakes. [3] According to the WHO (World Health Organisation) report 2024 (Fig 1), 150.2 million children under 5 years were stunted (too short for their age), 42.8 million were detected as wasted (too thin for their height), and 35.5 million were affected with over-weight (too heavy for their height). [4]

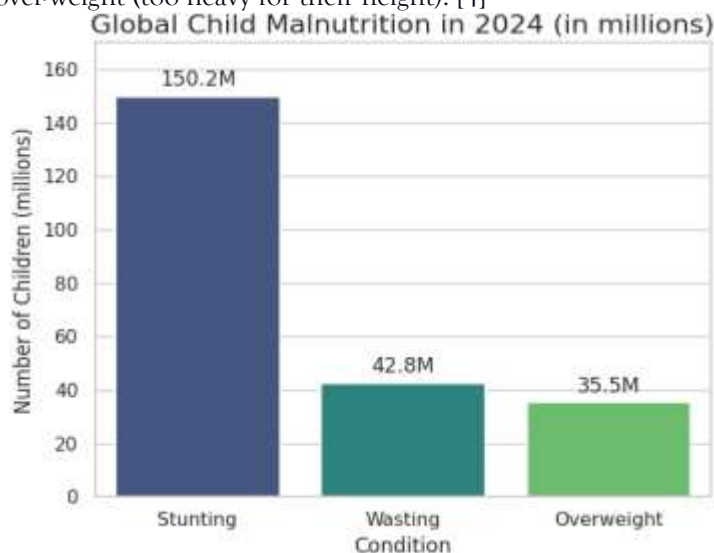


Fig. 1. Global Child Malnutrition Report [4]

Malnutrition problems are named as Triple Burden of Malnutrition (TBM), which includes undernutrition, overnutrition and micronutrient deficiency. TBM is recognised as the fastest-growing health challenge among children under five. Various studies from 1972 to 2022 identified multiple factors associated with malnutrition. Insufficient diet and illnesses are the main causes of the TBM. Lifestyle nutritional practices, contaminated water, insufficient food, lack of hygiene, improper feeding practices, poor health infrastructure and inadequate government health schemes implementations are the other subsequent reasons for malnutrition. Though various government schemes through Anganwadi centres and schools are in progress, TBM persists in children. Scientific evidence proves that the TBM affects the child's physical and mental development. The short-term consequences of malnutrition in childhood are fixed and irreversible damage in physical and cognitive growth, a weakened immune system, delayed development in physical and mental ability, frequent illness and micronutrient deficiency. The long-term consequences of malnutrition in adulthood are shorter height (linked with childhood stunting), risk of obesity (linked with childhood overweight/obese), lower earning potential, lower financial productivity, poor reproductive health, risk of chronic diseases and lower IQ. [5]

The multidimensional data were collected based on the child's characteristics, socio-economic, demographic, caring practices, maternal period hygiene, sanitation practices, health and immunisation practices. The study reveals that wasting condition is associated with the child's birth order, breastfeeding practices, and vaccinations. In the same way, the stunting condition is associated with age, father's occupation, parents' education status, breastfeeding status, mother's marriage age, diet routine during pregnancy and water and sanitation process. The underweight condition is related to the age, parents' education, child birth order, birth weight, vaccinations, diet routine during pregnancy, breastfeeding practices and water and sanitation process. [6] The Social Ecological Model (SEM) depicts the status of Severe Acute Malnutrition (SAM) in children. This model examines the factors associated with the influences of individual, relational, community, organisational and social level outcomes.

- The individual level analyses the biological factors, which include child clinical features like age, weight, disease or infections, medical history and nutritional status.
- The interpersonal level is related to the caregivers and home environment. It consumes family, social and household dynamics like parenting practices, family income and education status, hygiene and feeding practices and emotional support by the caregivers.
- The organisational level focuses on institutions related to service providers that offer care and treatment to children with SAM. It includes nutrition rehabilitation centres, health care centres, educational institutions, religious centres and local organisations.
- The community-level factors related to social behaviours. The neighbourhood conditions, cultural customs and community health workers of the broader community environment shape the care for the children.
- The public policy level includes health and social service schemes, government aid and food security. These policies either enhance or restrict community responses to SAM.[7]

The anthropometry measurements participate in the early detection of child growth and health issues through malnutrition detection. The positive outcome of nutrition level increases the academic performances and also emphasises the healthy weight practices for child health. The negative outcomes of nutritional status restrict the long-term socioeconomic benefits. [8] The maternal factors play a vital role in wasting conditions in toddlers. The factors like age, education, knowledge on parenting style, BMI, birth spacing and family income should be addressed to avoid wasting condition in children. Mother's education on nutrition, proper feeding method practices, child care practices and healthy pregnancy planning are the prominent efforts to prevent children from wasting. [9]

The feeding practices are highly correlated with the nutritional status of the child. If the feeding practices are not handled well, the consequences of this improper feeding cause significant impacts on the growth and development of the children. This study insists that the proper feeding behaviour of the caregivers can improve the nutritional status of the toddlers. [10]

The Spearman's rank correlation coefficient statistical data analysis was conducted to detect the relationship between the feeding practices of caregivers and the nutritional status of the toddler. The Spearman's rank test proves that feeding behaviour can enhance the child's nutrition status, with a value of. [10] In this study, the chi-square test was employed to find the most dominant features to identify undernutrition and malnutrition in toddlers. It took nutritional knowledge, diseases and eating habits as the predictor variables to predict malnutrition. The analysis reveals that the eating habits contribute 76%

to detecting malnutrition and undernutrition in toddlers. The other factors, nutritional knowledge and infectious diseases, affect the children with 15% and 19% respectively. [1] The statistical tools SPSS version 2.0 and WHO Anthro version 3.2.2 were used to analyse the factors associated with malnutrition of the children aged 6–59 months. The Chi-square test and Fisher’s exact test were implemented for this data analysis. The results show that undernutrition (12%) is the major health issue, while children with wasted status were 6.4%, stunted were 16.9% and overweight or obese were contributing 2.3%. [6]

Artificial intelligence (AI) and machine learning are redefining healthcare. It improves the diagnostics and medication methods. While the majority of the traditional health care services are manual, time-consuming, error-prone and manpower-intensive, the AI-driven health care services provide enhanced image analysis, personalised medicines, fast-tracks drug discovery, pattern detection, remote patient monitoring, nanotechnology, data management, robotic surgeries and predictive analysis. The AI-driven technology focuses on precision, speed and predictability. [11][12][13] The AI solutions were related to wearable devices to monitor patients’ abnormalities to provide early warnings.[14] The machine learning and deep learning techniques were employed to detect anaemia in children using conjunctiva images. The Convolution Neural Network (CNN) predicts anaemia with 98.45% accuracy. [15]

The physical, cognitive and emotional growth of the infants and toddlers strongly depends on their sleep, eating habits, medical attention, healthy behaviours and parents. High fibre, calcium, iron and protein are essential for healthy children. The machine learning algorithms facilitate understanding and predicting malnutrition in children. [18] The machine learning algorithms were implemented to predict the multiple malnutrition statuses: stunting, malnutrition, undernutrition, adequate nutrition, severe malnutrition, overnutrition and obesity. Inadequate nutrition in children can affect their cognitive abilities. It is important to address the nutritional status of the child as early as possible. Stunting causes a chronic nutritional problem in children, which affects their physical and cognitive abilities. [17][19][21][22] The data pre-processing technique Principal Component Analysis (PCA) was employed for feature selection, and a naïve Bayes classifier model classified the nutritional status of the toddlers into six categories: good nutrition, at risk of excess nutrition, excess nutrition, obesity, undernutrition, and severe malnutrition. It achieved 85.76% accuracy. [23] The detection of the factors which affect the accuracy of the random forest machine learning model in the process of toddler malnutrition classification. With 99.91% of accuracy, 100% of precision, 100% of recall and 100% F1-score, the factors – age, gender and height contributed strongly in the random forest model. [24]

The table below (Table II.) analyses the various machine learning algorithms' performances. It addresses the problems related to malnutrition in toddlers. The major malnutrition problems are stunting and wasting. It also identifies the features related to malnutrition issues in toddlers. The features related to anthropometry measurements and environmental conditions. The machine learning models take most of the input features as anthropometry measurements, age, height and weight.

Problem Statements	Input Features	Target Variable	Machine Learning Algorithms	Accuracy
Paediatric health diagnostics [16]	Age, Height	Height Condition	Logistic Regression	94.65%
Understanding and Predicting the Prevalence of Malnutrition [18]	Sleeping Routine, Poverty, Parents Knowledge, Food Allergies, Nutrition Quantity, Food Supplements, Unsanitary Conditions, Over Feeding, Promoting Regularity, Satiety, Calcium, Protein, Fruits and Vegetables.	Ecological Conditions, Family Care , Physical Health and Serving Size	Decision Tree	92%
			SVM	91%
			KNN	84%
			GBM	79%
			Random Forest	65%
Nutrition Classification [19]	Body Weight (BB), Height (TB), Age, BB/TB Ratio, Z-Scores for BB/U, Z-Scores For TB/U, and Z-Scores for BB/TB	Nutritional Status into Six Class Labels: malnutrition, undernutrition, adequate nutrition, overnutrition, risk of	SVM	89.04%

		overnutrition, and obesity		
Nutritional Status Classification [20]	Age, Body Mass Index (BMI), Weight, and Body Length	Nutritional Status Into Four Class Labels: Normal, overweight, underweight and malnutrition	Naïve Bayes	89%
			Linear Discriminant Analysis	92%
			Decision Tree	97%
			KNN	74%
			Random Forest	97%
			SVM	96%
Predicting Stunting Nutritional Status [22]	Age, Weight, Height, and Parental Nutritional Status	Stunting Status	Naïve Bayes Classifier	85%

TABLE II. Existing Machine Learning Models

The existing researches contain multiple limitations which control the overall effectiveness and scope of the proposed systems. The previous studies focus on small or low-dimensional datasets. It fails to extract the depth and heterogeneity of features contributing to the child's health related to malnutrition. Most of these studies leverage the traditional machine learning algorithms. There is a lack of advanced or hybrid approaches which can enhance the predictive model's accuracy and reliability. Another common drawback is the utilisation of single-dimensional features, for example, diet routine or anthropometric factors. The clinical and socio-economic aspects of features need to be addressed to predict the malnutrition health status of the children. At the point of target variable prediction, it often narrows down to the single perspective of malnutrition, either stunting or wasting. There is an absence in addressing all types of primary malnutrition in children, such as wasting, stunting, and underweight.

In this study, the holistic view of malnutrition health status will be detected by combining the status of wasting, stunting, and underweight in children up to 5 years. The multi-dimensional features set were used to predict the target variable. The prediction approach of the malnutrition health status of the children will utilise the advanced machine learning techniques with enhanced predictive accuracy and robustness.

MACHINE LEARNING ALGORITHMS

Logistic Regression Classifier: Logistic regression is one of the supervised learning techniques which solves problems through a linear approach. The logistic regression model trains with labelled data and predicts the outcomes for the unlabelled new data. It is applied for the categorical target variable for binary classification, but it can also be applied to the multi-level target variable. [37] The logit link function helps to detect the log odds of the outcome for the input variables. The logistic function provides a statistical way of transforming the log odds into the probability for the predictor variable. The relationship between the independent and dependent variables is non-linear in the case of probability theory, but it is linear in the case of log odds. The primary goal of the logistic regression model is to calculate the parameters through Maximum Likelihood Estimation (MLE). This MLE provides the group of variables that increases the probability of the input data. The regression coefficient provides the strength and direction of the relationship between the input and output variables. The logistic regression technique is one of the most commonly used techniques in various fields like healthcare, finance, marketing, etc. [38]

Decision Tree Classifier: A Decision tree classifier is one of the widely popular techniques in creating a classifier model. The decision tree is constructed with nodes and branches. Each node represents the categorical classes of the target attribute. [31] As it holds the tree-like structures, it can handle large datasets. The outer branch depicts the results, and the inner branch depicts the dataset properties. [32] The training of a decision tree begins from the root node, which holds the entire dataset. The homogeneous subgroups are divided by evaluating the feature split criteria information gain (based on entropy) or Gini impurity. The threshold and best feature are selected based on this evaluation process. The child nodes are created with this best feature, and this process continues recursively until the stopping criteria are met. The stopping conditions are the maximum depth of the tree, the minimum number of instances per leaf and when all instances in a node are classified into the same class. The classification

rules are formed by deriving every path from the root node to the leaf node. For new data points, the predictions are derived by traversing from the root node and based on the attribute values until arriving at a leaf node. [32] This makes the decision tree model fast and transparent in the classification process.

Random Forest Classifier: Random forest classification is one of the powerful supervised machine learning algorithms for classifying the input data into appropriate categories. Random forest classification is executed based on the bagging concept. The concept of bagging is used to improve the prediction capability by combining multiple decision trees. Each decision tree acts as an individual learner. They are trained independently with unique samples of data. The original dataset is divided into multiple samples of data based on the number of decision trees. The random sets of features are selected at the time of tree construction. [29] This reduces the correlation among the decision trees. Once the decision trees have been created, the random forest employs majority voting for the prediction process. All the trees trained with a randomly selected dataset grow differently. For the new data, each decision tree predicts one of the target classes. The most voted class will be selected for the outcome. Rather than deriving the result from one decision tree, this ensemble technique provides better generalisation on new or unseen data. [30] Because of its bagging concept, the overfitting issue can be raised. The random forest alleviates the overfitting risk by demonstrating the randomness in both data points and feature sets. It can handle large and multi-dimensional datasets. It enables the robustness and increases the accuracy of the model. The random forest classification is incredibly effective in many real-world scenarios, especially where the non-linear and complex relationships among the feature set are present.

Gradient Boosting Classifier: A Gradient boosting classifier is a supervised machine learning technique which utilises the ensemble learning method. This ensemble learning technique combines various weak learners to create a strong and extremely efficient predictive model. The technique behind this ensemble model is to integrate the predictions of various models and get an accurate result which is more robust than any individual model. In this case, the weak learners are usually the decision trees; each decision tree's mistakes are corrected sequentially to get more accurate results. [39] These decision trees individually capture distinct patterns of the dataset, which is then aggregated. Its ability to limit the overfitting issue enhances the predictive accuracy. [40]

XGBoost Classifier: The extreme gradient boosting classifier is a highly effective classification method. It is an end-to-end tree-based algorithm. It is a highly scalable boosting system. [33] The decision tree's training speed is increased through various methods. It finds the best split to reduce complexities in computations. The best split attributes are selected based on the highest gain value. The linear scanning process is performed on every sorted attribute. The XGBoost classifier simplifies the sorting process by using a structure called a compressed column, which holds pre-sorted attributes. Rather than scanning every possible feature split, the percentiles of the data are calculated. The gain is calculated through the aggregated statistics. [34] It combines the predictions of the multiple decision trees. The decision trees are organised sequentially. Each new tree is learned to minimise the previous tree's errors (loss function). The XGBoost classifier is one of the trusted machine learning algorithms in classification tasks by incorporating flexibility, speed and accuracy enhancement. The better result in a short span of time is achieved by combining both hardware and software optimisation approaches. The regularisation parameters are used to ignore the overfitting issue. [35]

Performance Metrics: The objective of this study is to classify the malnutrition health status of children under the age of five. Based on the multi-dimensional input data, the malnutrition health status of the children is classified into five categories or classes. This type of problem is classified as a classification problem. The classification problem is solved by making assumptions on the given categorical classes. The learned knowledge of the training process with training sets and class labels will be applied to the newly acquired data to predict the corresponding class label. [31] The performance of the proposed model is evaluated through the quantitative measurements such as accuracy, precision, recall, and F1-score and graphical representation of the learning curves. [36]

Accuracy (A) is referred to as the ratio of the correctly predicted results with the total observations.

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision (P) is the ratio between the correctly predicted positive instances and total predicted positive instances.

$$P = \frac{TP}{TP + FP}$$

Recall (R) is represented by the ratio of correctly predicted positive observations to all actual positive instances.

$$R = \frac{TP}{TP + FN}$$

F1-score is measured as the harmonic mean of precision (P) and recall (R).

$$F1\ score = 2 \times \frac{P \times R}{P + R}$$

PROPOSED METHODOLOGY

Data Preparation

The dataset was prepared in a synthesised manner to predict the malnutrition health status of the children between 0 and 5 years. The health indicators were selected based on the expert-paediatrician suggestions, and the target variables were derived with the help of the POSHAN tracker website, which is maintained by the Ministry of Women and Child Development, Government of India. According to the POSHAN Tracker reports, around 7 crore children (0-5 years) were assessed for nutritional growth. The finding indicates -

- 34% children were in Stunting (Severely/Moderately Stunted) status,
- 14% children were in Underweight (Severely/Moderately Underweight),
- 6% children were in an Overweight (Obese/Overweight) condition and
- 5% children were dealing with the Wasting (SAM/MAM) condition.

The NFHS-5 survey (2019-21) revealed the status of three types of malnutrition:

- Stunting condition includes 35.5% of the children,
- Underweight status includes 32.1% of the children and
- Wasting status includes 19.3% of the children

The synthesised dataset was prepared based on the above statistics and ensures that it reflects these trends.

The scoring method was applied to the malnutrition status attributes: Stunting_status, Underweight_status and Wasting_status to derive the target variable: health_status. Based on the severity combinations of the Stunting_status, Underweight_status, and Wasting_status attributes, the target variable was formed with four categories: Normal, Mild, Moderate and Severe.

Dataset Description

The dataset includes 21 attributes and 115724 records. There are four sets of input features that were synthesised, and the corresponding target variable was derived through a rule-based label engineering technique.

- Demographic Data: DOB and Gender were included since they directly affect the outcomes of the malnutrition health status of the children.
- Anthropometric Measurements: Height and Weight of the child.
- Clinical and Developmental Indicators: Birth Weight, Birth Order, History of Infections, Developmental Delays, Vaccination Status, Appetite and Energy Level and General Behaviour.
- Socio-Economic and Environmental Indicators: Mothers' Education Level, Household Monthly Income, Type of Residence, Anganwadi Centre Beneficiary, Meal Frequency and Healthcare Access. The detailed dataset description is shown in the table below.

Variable Name	Description	Data Type	Possible Values/Format
DOB	Date of birth of the child	Date	MM/DD/YYYY or DD-MM-YYYY
Gender	Gender of the child	Categorical	"Male", "Female"
Height	Height of the child (in cm)	Numeric	Continuous values (e.g., 97.2, 115.6)
Weight	Weight of the child (in kg)	Numeric	Continuous values (e.g., 11.897, 20.245)
Birth weight	Categorical birth weight	Categorical	"Below 2.5", "2.5 or Above" (kg)
Birth order	Order of birth among siblings	Categorical	"1st Born", "2nd Born", "3rd Born"
History of Infections	Whether child had infections	Categorical	"Yes", "No"

Developmental Delays	Presence of developmental delays	Categorical	"Yes", "No"
Vaccination Status	Vaccination completeness	Categorical	"Fully vaccinated", "Partially vaccinated"
Appetite and Energy Level	Qualitative assessment	Categorical	"Very poor", "Poor", "Average", "Good", "Excellent" with descriptions
General Mood and Behavior	Behavioral assessment	Categorical	"Normal", "Frequently irritable or tired"
Mothers Education Level	Mother's highest education	Categorical	"No formal education", "Primary school", "Secondary school", "Higher secondary", "Graduate or higher"
Household Monthly Income	Monthly household income	Categorical	Ranges like "Below ₹10,000", "₹10,001-₹15,000", etc.
Residence Type	Type of residential area	Categorical	"Urban", "Semi-urban", "Rural"
AWC Beneficiary	Whether Anganwadi centre beneficiary	Categorical	"yes", "no"
Meal Frequency	Daily meal frequency	Categorical	"1 meal per day", "2 meals", "3 meals", "4 or more meals"
Access to Healthcare	Healthcare accessibility	Categorical	"Yes, regular access", "No, limited access"
Stunting_status	Child's height-for-age	Categorical	"Severely Stunted", "Moderately Stunted", "Normal".
Underweight_status	Child's weight-for-age	Categorical	"Moderate Acute Malnutrition (MAM)", "Severe Acute Malnutrition (SAM)", "Normal, Over-weight", "Obese".
Wasting_status	Child's weight-for-height	Categorical	"Severely Underweight", "Moderately Underweight", "Normal".
health_status	Overall malnutrition health status of the child	Categorical	"Normal", "Mild", "Moderate", "Severe".

Data Pre-processing

Data pre-processing is one of the essential steps in developing a machine learning model. Since the raw data is not always ready for use in ML and analysis, the pre-processing transforms the raw data into clean and machine-readable data. This process enhances the accuracy and reliability of the ML model outcomes and its insights into the data. The following pre-processing steps have been applied to the raw data:

- Feature Engineering: Date or temporal transformation was applied to convert the raw date attribute - DOB into Age_in_Days, which enables it to be usable in ML models.

$$\text{Age_in_Days}_i = (\text{Current_Date} - \text{DOB}_i), \forall i \in [1, N]$$

Where,

Age_in_Days_i is the age of the i - th child in days.

DOB_i is the i - th child's Date of Birth.

Current_Date is pre-processing date.

N is the sum of the samples (children) in the dataset.

This conversion provides continuous values in this attribute (Age_in_Days), which represents children's age precisely and makes it suitable for training the model. The accurate relationships between age and nutritional health status of the child have been observed efficiently in this classification problem.

- Categorical Encoding and Target Encoding: The collected dataset contains multiple categorical variables which are in non-numeric forms. Since machine learning algorithms require numerical input data for modelling, the categorical features \mathbb{X}_j (target attribute excluded) were converted into numeric values through label encoding as follows:

$$\mathbb{X}_j = f_{le}(\mathbb{X}_j)$$

Where, f_{le} is the label encoding function. It maps each class label to a unique numeric value through a mapping rule:

$$f_{le}: \{Class_1, Class_2, Class_3, \dots\} \rightarrow \{0, 1, 2, \dots\}$$

Similarly, the target attribute y (health_status) also encoded into numerical format to preserve the distinct class labels without any artificial assumption or relationship.

$$y = f_{le}(y)$$

- **Removal of Attributes:** The attributes Stunting_status, Underweight_status and Wasting_status have been removed to cut off the overfitting issues. The machine learning models may memorise the possible outcomes based on these attributes since the target variable was derived from these attributes. After removing these attributes, the remaining 17 input features and one target variable were sent to the next stage of the machine learning pipeline.

Train-Test Split

This step is an important step to verify the model's reliability and efficiency on the current dataset, but also with the real-time data. The imbalances on minority target classes were resolved with the SMOTE technique, and the train-test split occurs with a 75:25 ratio, respectively.

Balance Dataset

Though the synthesised dataset reveals the real-time class distributions in the attributes - Stunting_status, Underweight_status and Wasting_status, the newly derived target variable possesses the class imbalance, which may provide an overfitting issue in ML models. This issue was overcome with the help of the Synthetic Minority Oversampling Technique (SMOTE) at the training phase. This technique generates the synthetic samples of the minority classes to overcome the class imbalance.

$$\bar{X}_{train}, \bar{y}_{train} = SMOTE(X_{train}, y_{train})$$

Where, $\bar{X}_{train}, \bar{y}_{train}$ are the augmented training data after the SMOTE and X_{train}, y_{train} are the original training data. The following figures show the original class distribution and augmented class distribution after the SMOTE technique.

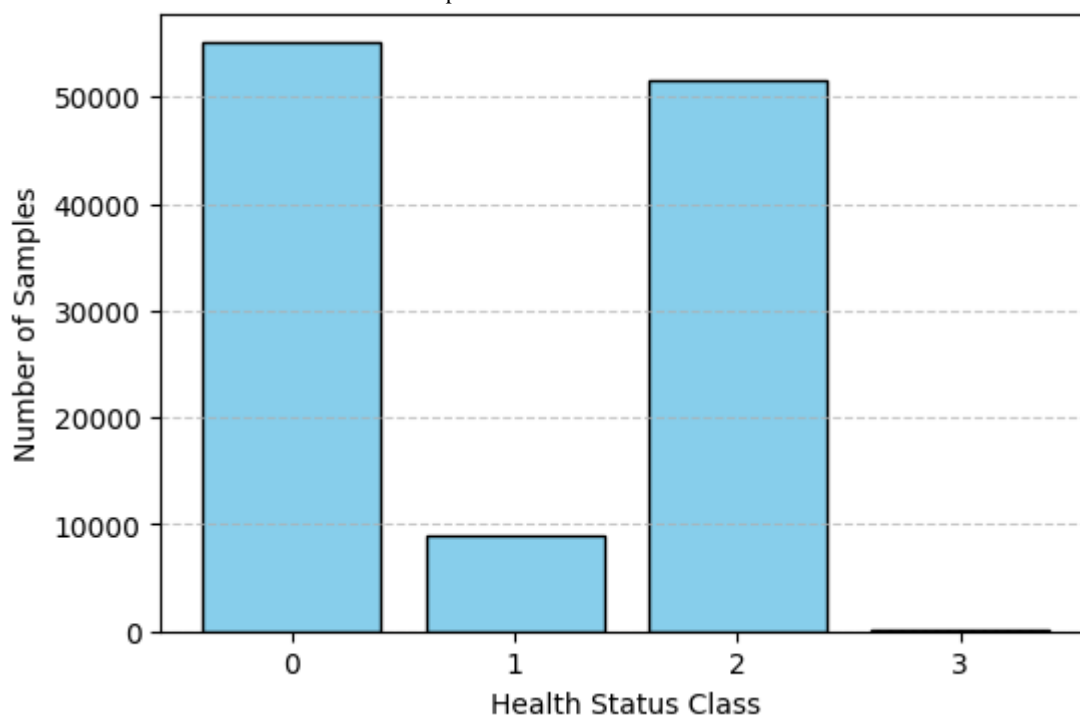


Fig. 2. Class Distribution Before Smote

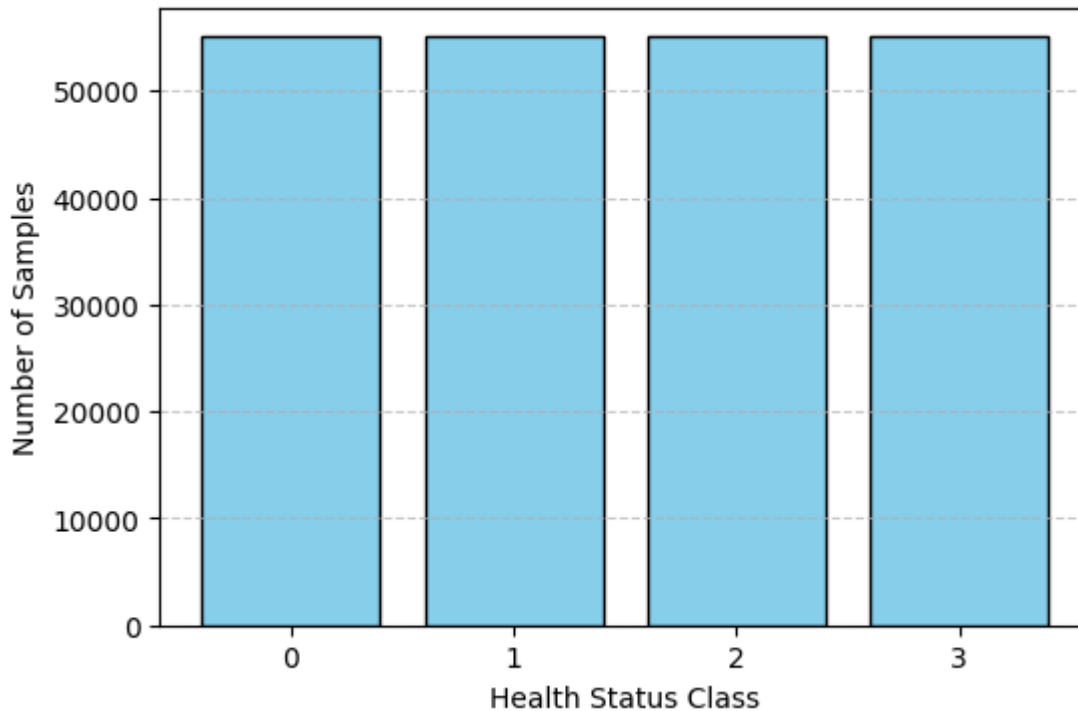


Fig. 3. Class Distribution After Smote

Hyperparameter Tuning

The performance of the machine learning models can be optimised through hyperparameter tuning. This study has applied the Grid Search method as a hyperparameter tuning process. The Grid Search technique evaluated every possible hyperparameter combination systematically in a predefined grid Θ . The goal is to detect the optimal hyperparameter configuration $\hat{\theta}$ for each model M with the set of classifiers $\{M_1, M_2, \dots\}$ and this enhances the cross-validation accuracy of the model.

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \frac{1}{k} \sum_{i=1}^k \text{Accuracy} \left(M(\theta), \left(\mathbb{X}_{train}^{(i)}, \mathcal{Y}_{train}^{(i)} \right) \right)$$

Where, k is the number of cross-validation (CV) folds, Θ is the hyperparameter grid which specifies all combinations of hyperparameters, $M(\theta)$ is the model trained with a specific hyperparameter configuration θ and $\left(\mathbb{X}_{train}^{(i)}, \mathcal{Y}_{train}^{(i)} \right)$ is the training data split in the i -th CV fold. This approach ensures that the ML model is not overfit to a specific data split and generalisation with the selected hyperparameters. When the hyperparameter configuration $\hat{\theta}$ is detected, the model is trained again with the training data using $\hat{\theta}$.

Train Best Model

Once the optimal parameters θ were identified, the final models M_1, M_2, \dots were trained using these parameters. The training process involved fitting the model on the balanced data set $\bar{\mathbb{X}}_{train}, \bar{\mathcal{Y}}_{train}$ which was synthesised through the SMOTE technique.

$$\hat{M} = M(\hat{\theta})$$

Where, M is the machine learning algorithm (Logistic Regression, Decision Tree, Random Forest, Gradient Boosting and XGBoost algorithms) and $\hat{\theta}$ are the optimal hyperparameters found during the hyperparameter tuning. The relationship between input features and target health_status attribute was learned by the model \hat{M} through the balanced dataset.

$$\hat{M} \leftarrow \bar{\mathbb{X}}_{train}, \bar{\mathcal{Y}}_{train}$$

Performance Evaluation

Once the training phase was done, the model's performances were evaluated through the testing process. The testing data \mathbb{X}_{test} were utilised to assess the model's generalizability and efficiency. The testing data contains only the input features, and it may have class imbalances.

$$\bar{\mathcal{Y}}_{test} = \hat{M}(\mathbb{X}_{test})$$

Where, $\bar{\mathcal{Y}}_{test}$ is the outcomes of the test data, \mathbb{X}_{test} input data for the testing process and \hat{M} is the trained model. In the same way, all five machine learning models were trained and tested. Once the models were tested, the performances of the models were evaluated through the classification metrics such as precision,

recall, F1-score and accuracy. These metrics help to understand how well the models predicted the malnutrition health status of the child (Normal, Mild, Moderate and Severe) on the unseen data.

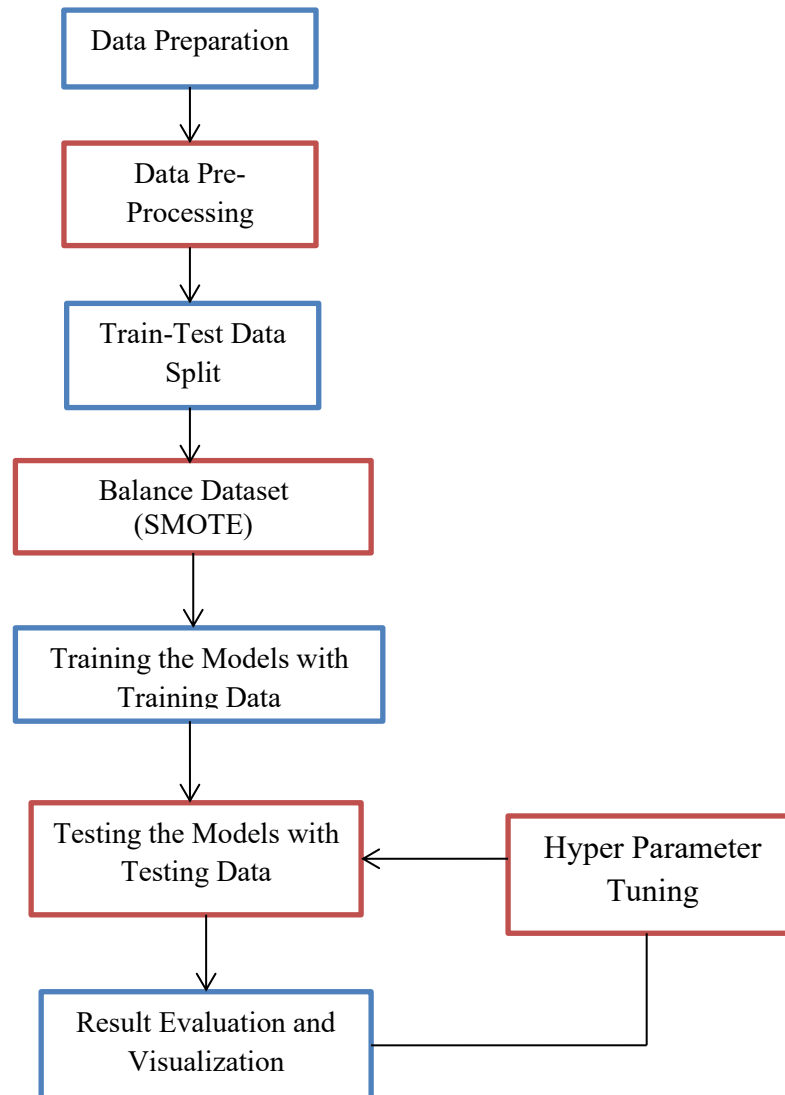


Fig. 4. Architectural Flow of the Proposed Methodology
Visualization of Learning Dynamics

The performances of the ML models were visualised through the confusion matrix and learning curves. These learning curves detect the overfitting and underfitting issues efficiently. For each size S , the training accuracy and validation accuracy were plotted, averaged over k -fold CV splits.

$$Training\ Accuracy(S) = \frac{1}{k} \sum_{i=1}^k Accuracy\left(\hat{M}, \left(\mathbb{X}_{train}^{(S,i)}, y_{train}^{(S,i)}\right)\right)$$

$$Validation\ Accuracy(S) = \frac{1}{k} \sum_{i=1}^k Accuracy\left(\hat{M}, \left(\mathbb{X}_{valid}^{(i)}, y_{valid}^{(i)}\right)\right)$$

RESULT AND DISCUSSIONS

This study has implemented various types of machine learning models to predict the overall child malnutrition health status dataset to build the predictive model. The classification algorithms, including linear model, Logistic Regression and non-linear model, Decision Tree algorithms were applied to detect the performance of the basic models. The ensemble techniques, such as bagging and boosting-based techniques, were applied to build the Random Forest machine learning model, Gradient Boosting and XGBoost machine learning models. The dataset was synthesised according to the real-time data. Since the final target variable was derived from the existing attributes through a rule-based technique, class

imbalance occurred in the target variable. This class imbalance issue was rectified through the Synthetic Minority Oversampling Technique (SMOTE). This approach was applied only to the training data set to avoid information leakage from the test data and to develop a robust model. The results of the five machine learning models were depicted in the table below (Table III).

ML Model	Accuracy	Class	Precision	Recall	F1-Score
Logistic Regression	0.8351	0	0.84	0.83	0.84
		1	0.94	0.86	0.90
		2	0.81	0.84	0.82
		3	0.74	0.36	0.48
Decision Tree	0.9838	0	0.98	0.98	0.98
		1	0.95	0.96	0.95
		2	0.99	0.99	0.99
		3	0.70	0.59	0.64
Random Forest	0.9626	0	0.97	0.95	0.96
		1	0.96	0.91	0.93
		2	0.95	0.99	0.97
		3	0.96	0.56	0.71
Gradient Boosting	0.9463	0	0.97	0.92	0.94
		1	0.96	0.92	0.94
		2	0.92	0.98	0.95
		3	0.74	0.51	0.61
XGBoost	0.9848	0	0.98	0.99	0.98
		1	0.97	0.95	0.96
		2	0.99	0.99	0.99
		3	0.83	0.64	0.72

TABLE III. Result Analysis before Hyperparameter Tuning

Among these five machine learning models, the XGBoost model showed the highest overall accuracy of 98% and the balanced F1-scores for each class. The decision tree model also performed very well, but it showed weak performance on identifying the minority class, and it reflects overfitting issues. All five of these models' results represent mild overfitting issues. The learning curves have been plotted to detect the overfitting issue more clearly.

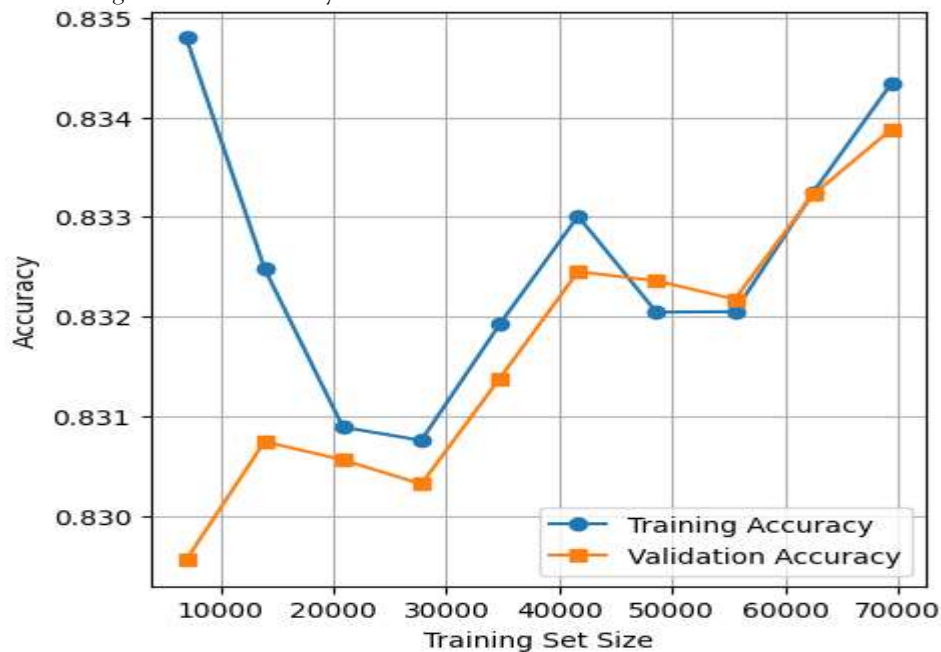


Fig. 5. Learning Curve of Logistic regression before Hyperparameter Tuning

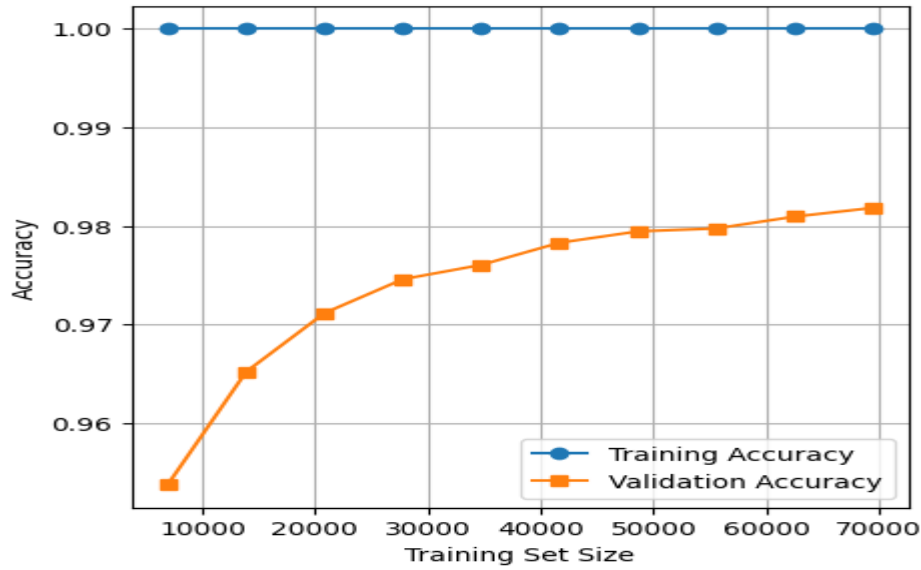


Fig. 6. Learning Curve of Decision tree before Hyperparameter Tuning

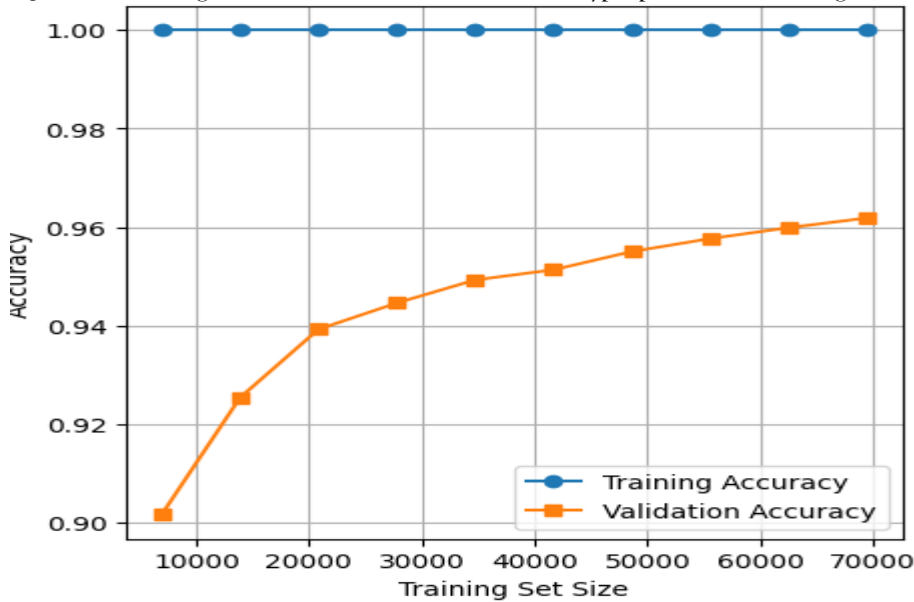


Fig. 7. Learning Curve of Random forest before Hyperparameter Tuning

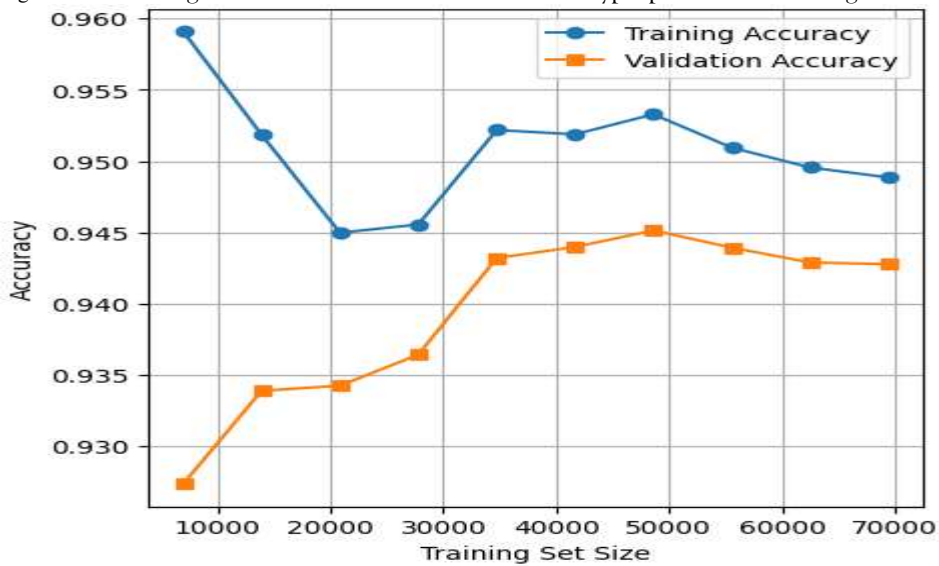


Fig. 8. Learning Curve of Gradient boosting before Hyperparameter Tuning

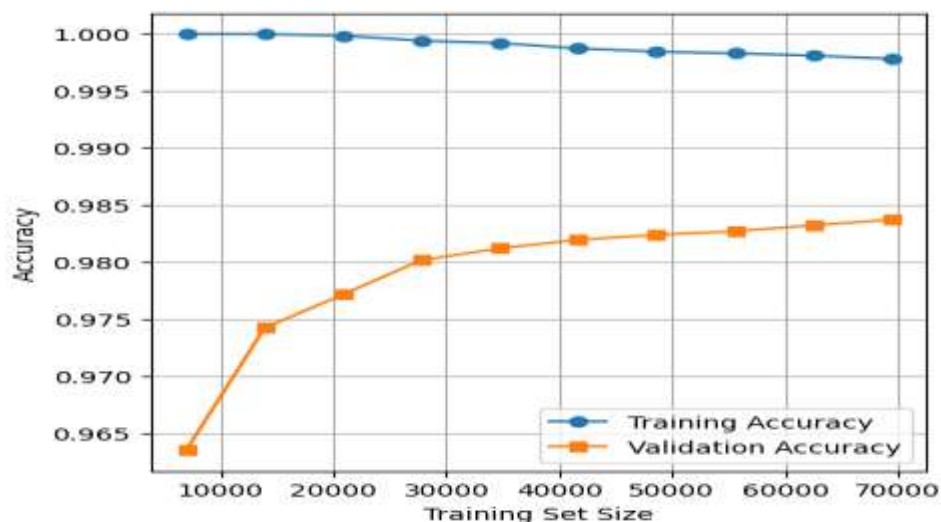


Fig. 9. Learning Curve of XGBoost before Hyperparameter Tuning

The learning curve analysis brings out the distinct behaviours of these machine learning models. The Decision Tree, Random Forest and XGBoost models exhibited extreme overfitting since the training accuracy was near 100% while the validation accuracy was much lower. The memorisation occurred rather than the generalisation of the new data set. The other two models, Gradient Boosting and Logistic Regression models, achieved near-optimal generalisation with the validation process. This overfitting issue was resolved with the help of a hyperparameter tuning process. The Grid Search CV method was implemented to find the optimal parameters for each ML model. The machine learning models were again retrained with the optimal hyperparameters to get better accuracy values without any overfitting.

ML Model	Accuracy	Class	Precision	Recall	F1-Score
Logistic Regression	0.8540	0	0.88	0.83	0.85
		1	0.62	0.91	0.74
		2	0.91	0.87	0.89
		3	0.11	0.79	0.19
Decision Tree	0.9799	0	0.98	0.98	0.98
		1	0.91	0.97	0.94
		2	0.99	0.98	0.99
		3	0.65	0.72	0.68
Random Forest	0.9633	0	0.98	0.94	0.96
		1	0.90	0.95	0.93
		2	0.96	0.99	0.97
		3	0.76	0.72	0.74
Gradient Boosting	0.9569	0	0.98	0.93	0.96
		1	0.86	0.94	0.90
		2	0.95	0.99	0.97
		3	0.56	0.90	0.69
XGBoost	0.9848	0	0.98	0.99	0.98
		1	0.97	0.95	0.96
		2	0.99	0.99	0.99
		3	0.83	0.64	0.72

TABLE IV. Result Analysis after Hyperparameter Tuning

The above table (Table IV) discusses the performance metrics of the five ML models after the hyperparameter tuning. The XGBoost machine learning model provided an overall accuracy of 98.4%. Followed by the Decision Tree, Random Forest, Gradient Boosting and Logistic Regression models achieved 97.9%, 96.3%, 95.6% and 85.4% as overall accuracies. The overfitting issue was analysed again based on the learning curve analysis. The learning curves of the enhanced machine learning models plot the curve of the training accuracy against the validation accuracy.

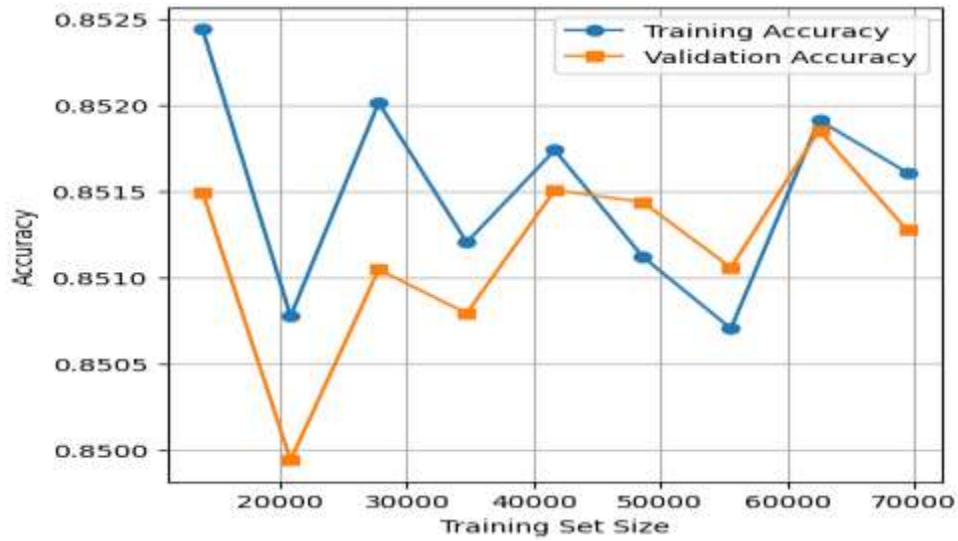


Fig. 10. Learning Curve of Logistic regression after hyperparameter tuning

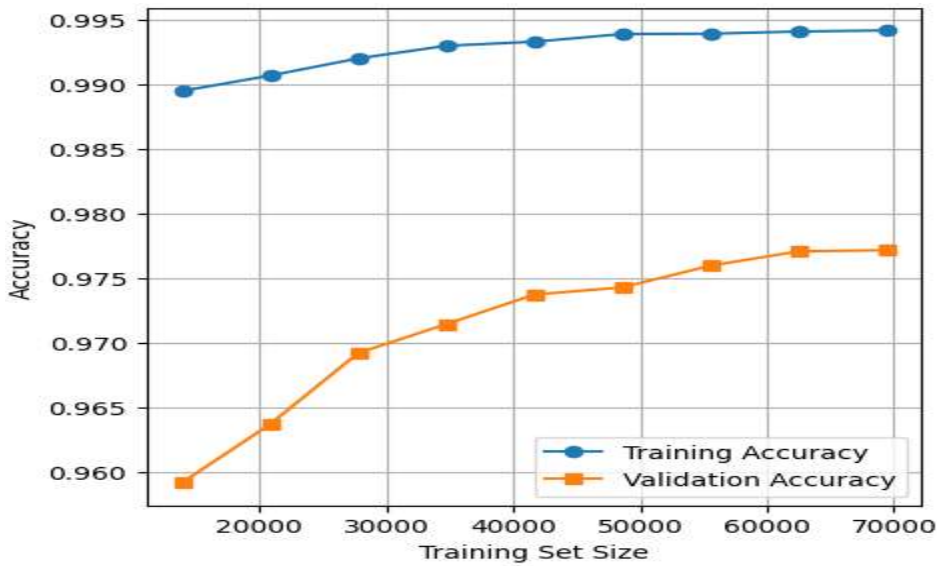


Fig. 11. Learning Curve of Decision tree after hyperparameter tuning

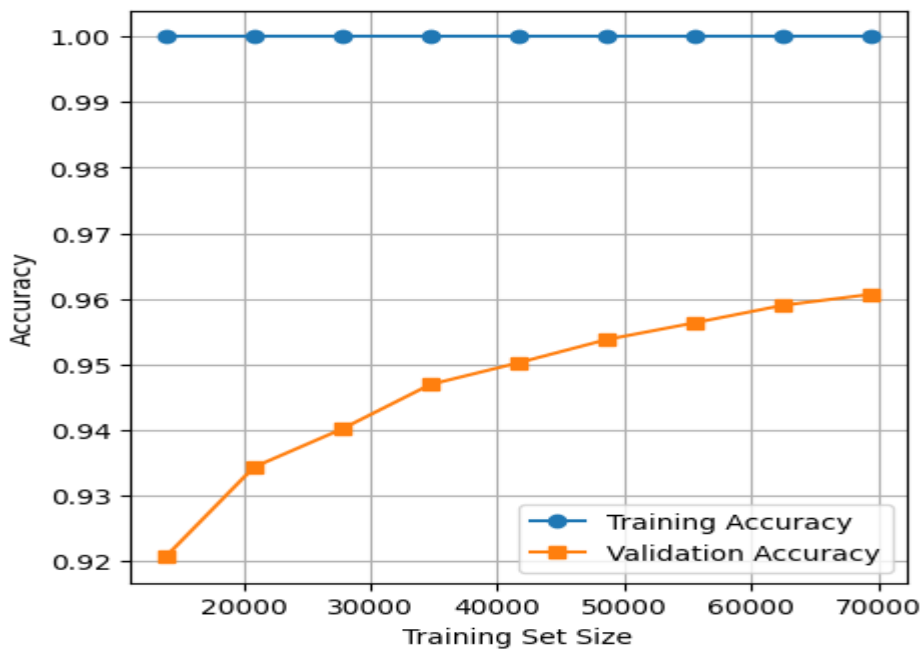


Fig. 12. Learning Curve of Random forest after hyperparameter tuning

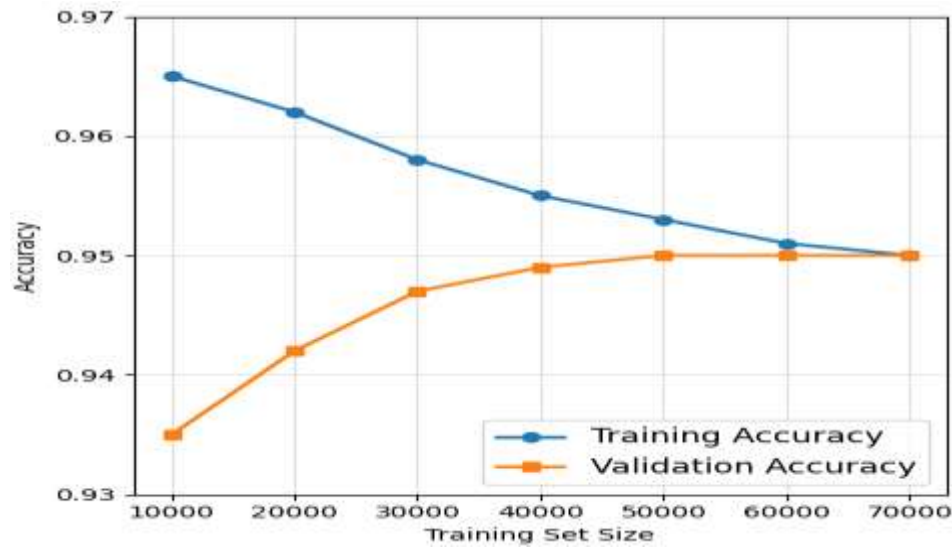


Fig. 13. Learning Curve of Gradient boosting after hyperparameter tuning

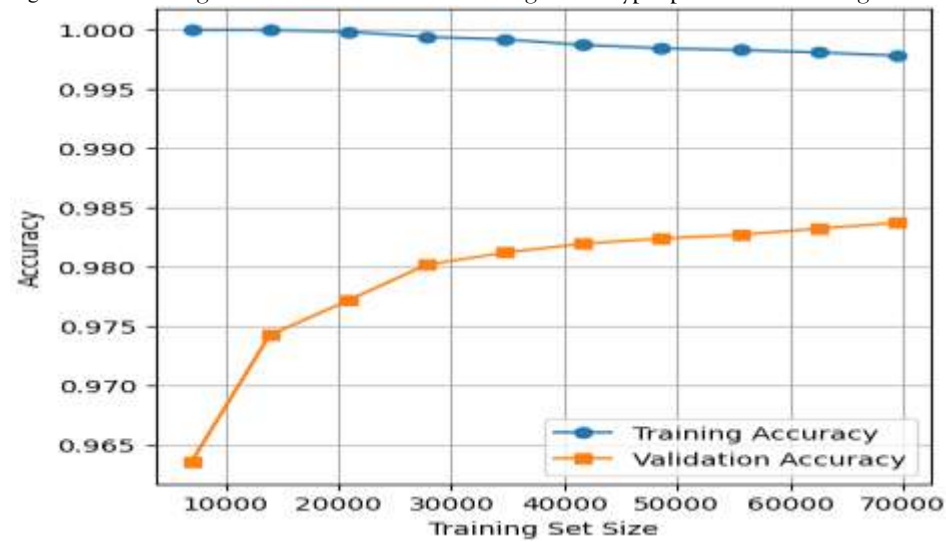


Fig. 14. Learning Curve of XGBoost after hyper parameter tuning

Among these learning curves of the enhanced machine learning models, only the Gradient Boosting machine learning model overcomes the overfitting issue, and at the same time, it has achieved the highest accuracy. Though the Logistic Regression model overcame this overfitting issue, the accuracy was not better than the gradient boosting model. The other three models still faced the overfitting issue, but the accuracies were better than the gradient boosting model. The confusion matrix of the Gradient Boosting model is shown in Fig 15. It demonstrates the strongest performance on classes 0, 1 and 2. Class 4 prediction was still challenging in this proposed model.

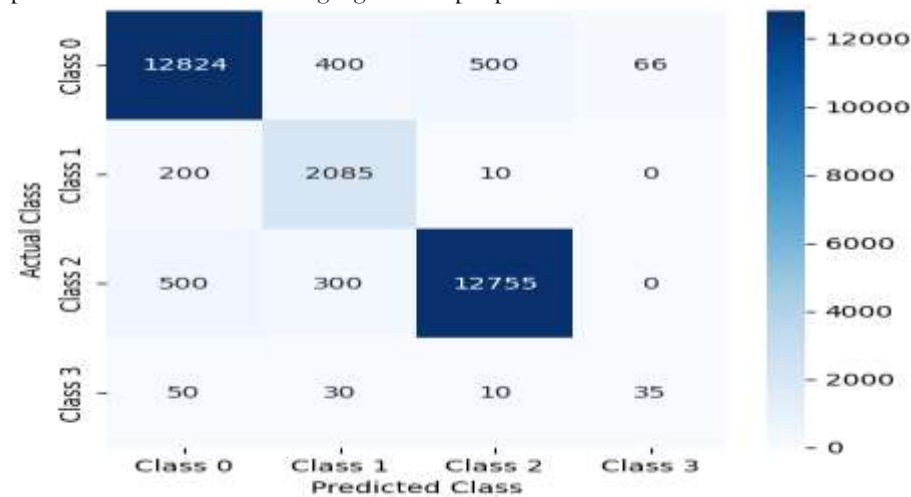


Fig. 15. Confusion matrix of Gradient boosting

The proposed model utilised two most impactful techniques – SMOTE to balance the dataset and Grid Search hyperparameter tuning to detect the optimal parameters for the machine learning models. The SMOTE technique improved the recall for the minority class label 3. Still, it doesn't reflect the proper distribution of the minority classes. The overfitting issue and poor generalisation should be addressed with an additional technique for robust evaluation and generalisation. The hyperparameter tuning process played a crucial role in the performance optimisation of the proposed machine learning model. The class imbalances and weak generalisation are not only handled by the hyperparameter tuning process. Additional improvements and techniques may overcome these issues.

CONCLUSION

This study intends to demonstrate the effectiveness of the machine learning system in the prediction of malnutrition health status among children under five years. The proposed machine learning model analysed relationships among the multi-dimensional factors - anthropometry measurements, clinical and developmental attributes, socio-economic and environmental indicators. As this malnutrition detection problem consists of a categorical target variable, the classification machine learning algorithms Logistic Regression, Decision Tree, Random Forest, Gradient Boosting and XGBoost were employed to construct the machine learning models. The relationship pattern among these multi-dimensional features and the target variable – malnutrition health status has been derived from this proposed machine learning model, Gradient Boosting, with 95.6% of accuracy. The SMOTE technique was applied to rectify the class imbalances and enhance the prediction accuracy of the model. The overfitting issue was solved by the hyperparameter tuning. This research can be helpful for both officials and common people to identify the risk of malnutrition and the health status of the children at an early stage. This makes them take prevention measures and intervention programs for targeted populations, as well as the individuals.

FUTURE ENHANCEMENT

The future research may examine the advanced paradigms, such as hybrid models or quantum machine learning, to overcome the complexity of the multi-dimensional dataset. It also enhances the scalability, privacy and computational efficiency in early malnutrition detection. These innovative techniques could unlock faster and more accurate assessments of the malnutrition health status of the children, even in a huge and diverse set of populations.

REFERENCES

- [1] Ratnawati, Mamik, Monika Sawitri Prihatini, Ririn Probowati, Rini Hayu Lestari, and Dwiharini Puspitaningsih. "Analysis of The Situation of Undernutrition and Malnutrition on Toddlers." *Malaysian Journal of Medical Research (MJMR)* 6, no. 3 (2022): 1-5.
- [2] Reusken, Meike, Christopher Coffey, Frans Cruijssen, Bertrand Melenberg, and Cascha van Wanrooij. "Identification of factors associated with acute malnutrition in children under 5 years and forecasting future prevalence: assessing the potential of statistical and machine learning methods." *BMJ Public Health* 3, no. 1 (2025).
- [3] Padhani, Zahra Ali, Jai K. Das, Saeed Akhtar, Tariq Ismail, and Zulfiqar A. Bhutta. "Tackling protein-calorie malnutrition during world crises." *Annals of Nutrition and Metabolism* 78, no. Suppl. 1 (2022): 27-38.
- [4] <https://www.who.int/data/gho/data/themes/topics/joint-child-malnutrition-estimates-unicef-who-wb>
- [5] Chopra, Hariyansh, Bobby Paul, Amrit Virk, Gargi Pandey, and Chandrakant Lahariya. "Triple burden of malnutrition among children in India: Current scenario and the way forward." *Indian Journal of Pediatrics* 90, no. Suppl 1 (2023): 95-103.
- [6] Bhandari, Tripura. "FACTORS ASSOCIATED WITH NUTRITIONAL STATUS OF 6-59 MONTHS CHILDREN IN DHARAN SUB-METROPOLITAN CITY, SUNSAR." PhD diss., Department of Nutrition and Dietetics Central Campus of Technology Institute of Science and Technology Tribhuvan University, Nepal 2021, 2021.
- [7] Braxton, Morgan E. *Understanding the Relationship Between Key Clinical Variables of Childhood Malnutrition and Time-To-Recovery In Guatemala*. East Carolina University, 2022.
- [8] Abdullahi, Mohammed Danjuma, John Emenike, and Ahmed Sani Doma. "Anthropometric measurements and nutritional status for sustainable education and development of children in the north central zone." *Journal of Advanced Education and Sciences* 4, no. 2 (2024): 28-38.
- [9] Afrah, Rahmayani, Desmawati Desmawati, and Roza Sriyanti. "Tackling Toddler Malnutrition: Exploring Maternal Influences on Wasting".
- [10] Septyani, Ima Aulia, Umi Faridah, and Muhamad Jauhar. "Feeding Behavior and Nutritional Status of Toddlers in Rural Areas." *Indonesian Journal of Health Research and Development* 3, no. 1 (2025): 33-40.
- [11] Roy, Sudipta, Tanushree Meena, and Se-Jung Lim. "Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine." *Diagnostics* 12, no. 10 (2022): 2549.
- [12] Tariq, Muhammad Usman. "The transformation of healthcare through AI-driven diagnostics." In *Enhancing medical imaging with emerging technologies*, pp. 250-264. IGI Global, 2024.

- [13] Chen, Xiang. "Ai in healthcare: Revolutionizing diagnosis and treatment through machine learning." (2024).
- [14] Sabry, Farida, Tamer Eltaras, Wadha Labda, Khawla Alzoubi, and Qutaibah Malluhi. "Machine learning for healthcare wearable devices: the big picture." *Journal of Healthcare Engineering* 2022, no. 1 (2022): 4653923.
- [15] Asare, Justice Williams, William Leslie Brown-Acquaye, Martin Mabeifam Ujakpa, Emmanuel Freeman, and Peter Appiahene. "Application of machine learning approach for iron deficiency anaemia detection in children using conjunctiva images." *Informatics in medicine unlocked* 45 (2024): 101451.
- [16] Rodriguez-Marin, Mauro, and Luis Gustavo Orozco-Alatorre. "Advancing Pediatric Growth Assessment with Machine Learning: Overcoming Challenges in Early Diagnosis and Monitoring." (2025).
- [17] Pang, Xueqin, Christopher B. Forrest, Félíce Lê-Scherban, and Aaron J. Masino. "Prediction of early childhood obesity with machine learning and electronic health record data." *International journal of medical informatics* 150 (2021): 104454.
- [18] Zumma, Thoufiq, Anikur Rahaman, Nuzhat Noor Islam Prova, Tasriful Haque, Julius Caesar Joy Bose, and Ravina Akter Youki. "Early Detection of Childhood Malnutrition Using Survey Data and Machine Learning Approaches." In *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, pp. 833-838. IEEE, 2025.
- [19] Asyraffi, Affan, Okfalisa Okfalisa, Fitri Insani, Surya Agustian, and Riski Mai Candra. "Toddler nutritional status identification: Support vector machine (SVM) algorithm adoption." *Science, Technology, and Communication Journal* 5, no. 2 (2025): 27-36.
- [20] Gustriansyah, Rendra, Nazori Suhandi, Shinta Puspasari, and Ahmad Sanmorino. "Machine learning method to predict the toddlers' nutritional status." *Jurnal Infotel* 16, no. 1 (2024): 32-43.
- [21] Janawisuta, Herianto, and Putu Harry Gunawan. "Early detection of stunting in Indonesian toddlers: A machine learning approach." In *2024 International Conference on Data Science and Its Applications (ICoDSA)*, pp. 12-16. IEEE, 2024.
- [22] Hariyanto, Rudi, Mohammad Zoqi Sarwani, and Yunita Nur Aprilia. "Prediction of Stunting Nutritional Status in Toddlers Using Naive Bayes Classifier Algorithm." *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)* 10, no. 2 (2025): 1111-1117.
- [23] Kamil, Muhammad Insan, and Adityo Permana Wibowo. "Implementation of the Naive Bayes Classifier Algorithm for Classifying Toddler Nutritional Status." *Journal of Applied Informatics and Computing* 8, no. 2 (2024): 421-427.
- [24] Fatmawati, Maylia, Bambang Agus Herlambang, and Noora Qotrun Nada. "Random Forest Algorithm for Toddler Nutritional Status Classification Website." *Journal of Applied Informatics and Computing* 8, no. 2 (2024): 428-433.
- [25] Morales, Fátima, Sergio Montserrat-De la Paz, Maria J. Leon, and Fernando Rivero-Pino. "Effects of malnutrition on the immune system and infection and the role of nutritional strategies regarding improvements in children's health status: A literature review." *Nutrients* 16, no. 1 (2023): 1.
- [26] Keats, Emily C., Jai K. Das, Rehana A. Salam, Zohra S. Lassi, Aamer Imdad, Robert E. Black, and Zulfiqar A. Bhutta. "Effective interventions to address maternal and child malnutrition: an update of the evidence." *The Lancet Child & Adolescent Health* 5, no. 5 (2021): 367-384.
- [27] Strielkowski, Wadim, Andrey Vlasov, Kirill Selivanov, Konstantin Muraviev, and Vadim Shakhnov. "Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review." *Energies* 16, no. 10 (2023): 4025.
- [28] Jayatilake, Senerath Mudalige Don Alexis Chinthaka, and Gamage Upeksha Ganegoda. "Involvement of machine learning tools in healthcare decision making." *Journal of healthcare engineering* 2021, no. 1 (2021): 6679512.
- [29] Rani, Pooja, Rajneesh Kumar, Nada MO Sid Ahmed, and Anurag Jain. "A decision support system for heart disease prediction based upon machine learning." *Journal of Reliable Intelligent Environments* 7, no. 3 (2021): 263-275.
- [30] Song, Minseok, Hyeyoom Jung, Seungyong Lee, Donghyeon Kim, and Minkyu Ahn. "Diagnostic classification and biomarker identification of Alzheimer's disease with random forest algorithm." *Brain sciences* 11, no. 4 (2021): 453.
- [31] Charbuty, Bahzad, and Adnan Abdulazeez. "Classification based on decision tree algorithm for machine learning." *Journal of applied science and technology trends* 2, no. 01 (2021): 20-28.
- [32] Bhatt, Chintan M., Parth Patel, Tarang Ghetia, and Pier Luigi Mazzeo. "Effective heart disease prediction using machine learning techniques." *Algorithms* 16, no. 2 (2023): 88.
- [33] Liew, Xin Yu, Nazia Hameed, and Jeremie Clos. "An investigation of XGBoost-based algorithm for breast cancer classification." *Machine Learning with Applications* 6 (2021): 100154.
- [34] FBentéjac, Candice, Anna Csörgő, and Gonzalo Martínez-Muñoz. "A comparative analysis of gradient boosting algorithms." *Artificial Intelligence Review* 54 (2021): 1937-1967.
- [35] Gupta, Ketan, Nasmin Jiwani, and Neda Afreen. "Liver disease prediction using machine learning classification techniques." In *2022 IEEE 11th International conference on communication systems and network technologies (CSNT)*, pp. 221-226. IEEE, 2022.
- [36] Liang, Haobang, Jiao Li, Hejun Wu, Li Li, Xinrui Zhou, and Xinhua Jiang. "Mammographic classification of breast cancer microcalcifications through extreme gradient boosting." *Electronics* 11, no. 15 (2022): 2435.
- [37] Acito, F. (2023). Logistic Regression. In: Predictive Analytics with KNIME. Springer, Cham. https://doi.org/10.1007/978-3-031-45630-5_7
- [38] Dey, Devjit, Md Samio Haque, Md Mojahedul Islam, Umme Iffat Aishi, Sajida Sultana Shammy, Md Sabbir Ahmed Mayen, Syed Toukir Ahmed Noor, and Md Jamal Uddin. "The proper application of logistic regression model in complex survey data: a systematic review." *BMC Medical Research Methodology* 25, no. 1 (2025): 15.
- [39] Singh, Priya, Swamy Gupta, and Vasu Gupta. "Multi-objective hyperparameter optimization on gradient-boosting for breast cancer detection." *International Journal of System Assurance Engineering and Management* 15, no. 5 (2024): 1676-1686.
- [40] Abbas, Farkhanda, Zhihua Cai, Muhammad Shoaib, Javed Iqbal, Muhammad Ismail, A. R. I. F. Ullah, Abdulwahed Fahad Alrefaei, and Mohammed Fahad Albeshr. "Uncertainty Analysis of Predictive Models for Water Quality Index: Comparative Analysis of XGBoost, Random Forest, SVM, KNN, Gradient Boosting, and Decision Tree Algorithms." (2024).