Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.6, December 2025, pp. 3312~3321
e-ISSN: 2548-6861

Classification For Determining Nutritional Status of Toddlers Using Random Forest Method at Tanah Pasir Primary Health Centre, North Aceh

Indana Zulfa [1]*, Mukti Qamal [2]*, Ar Razi [3]*
*Departemen of Informatics, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia
indana.210170014@mhs.unimal.ac.id [1], mukti.qamal@unimal.ac.id [2], ar.razi@unimal.ac.id [3]

Article Info

Article history:

Received 2025-08-20
Revised 2025-11-28
Accepted 2025-12-06

Keyword:

Toddler Nutrition Status,
Random Forest,
Data Mining,
Classification.

ABSTRACT

The nutritional status of toddlers is a fundamental factor in supporting their growth and development, particularly during the golden period of 0–5 years of age. Malnutrition in toddlers can have detrimental effects on physical growth, cognitive development, and immune function. In Indonesia, child malnutrition remains a significant public health challenge, particularly in rural areas, necessitating improved nutritional surveillance systems at primary health centers. The manual assessment of nutritional status at community health centers (Puskesmas) often poses challenges in promptly identifying toddlers with undernutrition or severe malnutrition. This study aims to develop a toddler nutritional status classification system based on the Random Forest method to assist healthcare workers in determining nutritional status quickly and accurately. This study utilized a dataset of 2,612 toddler anthropometric records collected from Tanah Pasir Community Health Center, North Aceh, between November 2024 and January 2025. The dataset was split into training (2,090 records, 80%) and testing (522 records, 20%) sets using stratified random sampling. Key variables included age (0-60 months), body weight (kg), and body height (cm). Nutritional status categories were determined based on WHO Child Growth Standards using the weight-for-age (W/A), height-for-age (H/A), and weight-for-height (W/H) indices. The Random Forest method was chosen due to its ability to construct multiple decision trees through ensemble learning, resulting in more accurate predictions and better resistance to overfitting. The model was implemented with 100 trees and evaluated using standard classification metrics. The experimental results demonstrated that the system achieved strong classification performance, with an accuracy of 93%, precision of 95%, recall of 98%, and an F1-score of 96%. The high recall value is particularly significant in healthcare applications, ensuring minimal false negatives in detecting malnourished toddlers. The developed system facilitates healthcare workers in efficiently and systematically monitoring toddlers' nutritional status with consistent classification standards. Therefore, this system is expected to serve as a decision-support tool to improve community nutritional status at the community health center level, enabling early intervention for at-risk children.

I. Introduction

Nutritional health for toddlers is a crucial aspect that requires careful attention, particularly from parents and healthcare professionals. Nutritional status reflects the level of success in meeting children's nutritional needs, which can be determined through indicators of body weight and height. This information serves as the foundation for assessing children's nutritional health conditions. Nutritional status plays an important role in achieving optimal health levels. Malnutrition can result in stunted growth, decreased energy for physical activities, weakened immune systems, and permanent brain function disorders, especially during the growth period of toddlers aged 0-5 years. Conversely, overnutrition also negatively impacts health, particularly causing overweight or obesity which can lead to various health risks [1].

Severe malnutrition or undernutrition is a problem frequently experienced by toddler groups and impacts health, growth, development, and productivity in adulthood. In assessing nutritional status, anthropometric measurement becomes a common approach by considering indicators of body weight, age, height, and gender of toddlers. Anthropometric measurement consists of a series of general indicators, including weight-for-age ratio (W/A), weight-for-height ratio (W/H), and height-for-age ratio (H/A). Generally, the W/A ratio is used as a standard method for assessing child growth, while the W/H ratio is considered the most sensitive anthropometric indicator in describing current nutritional status. The growth and health of toddlers play important roles in the progress of society and the nation. Therefore, collaboration from all parties is needed to fulfill shared responsibility in ensuring future generations grow well. Providing balanced nutritional intake appropriate to developmental stages is crucial for supporting optimal growth and forming healthy lifestyle habits in the future.

To understand the problem of malnutrition in toddlers in Indonesia, it is necessary to know that this condition is influenced by many factors, particularly food consumption patterns. The key contributing factor is limited parental nutritional knowledge, which is often influenced by low education levels and poverty. Additionally, prevention efforts must consider unhealthy environmental factors that exacerbate malnutrition risks. The importance of increasing public awareness about environmental hygiene, especially in certain urban areas, becomes crucial because dirty environments increase disease vulnerability in toddlers. It should be understood that toddlers who frequently fall ill tend to experience nutritional disorders. The serious impacts of this condition include physical growth disorders and mental development issues, which in turn can affect children's learning abilities and cognitive development in the future [2].

Understanding the importance of the toddler period as the golden phase of development, it should be recognized that this period determines a person's quality of life in the future. Child growth is not only measured by physical changes, but also reflects nutritional adequacy. To optimize growth and development, adequate nutritional fulfillment is absolutely necessary, while malnutrition risks disrupting development. In Indonesia, serious efforts are needed through regional mapping to identify areas with undernutrition cases. Strategic steps to improve the quality of nutritional intake are very important to ensure that toddlers' nutritional needs are met and to prevent the adverse effects of malnutrition [3].

Based on the background outlined above, a title can be proposed: "Classification for Determining Nutritional Status of Toddlers Using Random Forest at Tanah Pasir Primary Health Center, North Aceh". It is hoped that this research can identify malnutrition in toddlers.

II. METHODS

This research was conducted at Tanah Pasir Primary Health Center, North Aceh Regency. The selection of this location aims to facilitate the data collection process and references needed to support system design so that it can function optimally. The research was planned to take place from November 2024 to an undetermined time, starting from the research proposal preparation stage to the final research report preparation.

A. Research Workflow

The following are several stages of steps in conducting research:

1. Problem Analysis: Analyzing problems related to determining nutritional status in toddlers.
2. Data Collection: Collecting data by obtaining or requesting data from Tanah Pasir District Primary Health Center and conducting interviews with Tanah Pasir Primary Health Center staff regarding information needed in this research.
3. Data Processing: Processing data obtained from Tanah Pasir District Primary Health Center using Microsoft Excel software as a data processing medium.
4. Method Application: In this research, the author applies data mining techniques based on classification using the Random Forest algorithm to obtain solutions to the problems studied.
5. Data Testing: This stage is the process of testing the system that has been designed previously.
6. Conclusion: When all processes are completed, it will produce a conclusion obtained from the classification of toddler nutritional status determination.

B. Data Mining Method

Data is a collection of facts or information that can be measured, calculated, or processed. Data can be in the form of numbers, text, images, or other formats. In the field of computing, data is a digital representation of information that can be stored, processed, and transmitted by computer systems [4]. The data exploration process involves in-depth analysis of data archives to reveal relationships, certain patterns, or regularities in large datasets. The insights generated from this analysis play an important role in improving decision-making processes [5].

As part of Knowledge Discovery in Databases, data mining aims to transform raw data into knowledge through identification of significant patterns in historical data archives [6]. This process enables organizations to make more accurate data-driven decisions. Methodologically, data mining implementation integrates various disciplines including statistical analysis, mathematical models, intelligent systems, and machine learning techniques to filter crucial information from the ocean of data [7].

The data mining process can be divided into several interrelated and interactive stages [8]:

1. Data Cleaning: The dataset cleaning process aims to remove invalid, inaccurate, or unnecessary data. Meanwhile, data consolidation (data integration) involves collecting data from various sources into one integrated database.
2. Data Selection: To make analysis more efficient and precise, only relevant and necessary data will be selected from the database, considering that not all stored information will be used.
3. Data Transformation: To be analyzed effectively in data mining, data needs to be converted or unified into a compatible and standardized format.
4. Mining Process: This process becomes the heart of the activity where various algorithms are implemented to extract potential knowledge contained in the data. Several techniques are available for use, adapted to the type of analysis in data exploration.
5. Pattern Evaluation: Its main function is to detect meaningful relationships in data for further integration into the formed knowledge repository.
6. Knowledge Presentation: This process involves graphical representation and insight delivery about the approaches used, so that analysis results can be presented to users in an intuitive and easily digestible format.

C. Classification Method

One of the most important stages in data processing is classification [9]. Classification is the process of grouping or labeling new data or objects based on certain characteristics. Classification techniques are performed by analyzing variables generated from available data. The core of the categorization process is to determine labels or groups of unidentified entities. Stages in the classification system include model construction, model application, and evaluation process [10]. Essentially, this technique seeks to find a mathematical model capable of characterizing and separating various data classes, which can then be applied to predict classification of new data [11].

At this stage, classification of toddler nutritional status is performed through the application of the Random Forest algorithm. The advantages of this method include the ability to maintain accuracy despite missing values, robustness against extreme data, and optimization of storage space usage. Additional features in the form of attribute selection mechanisms help identify the most significant variables, which ultimately improve the quality of the classification model [12].

D. Random Forest Method

Random Forest is a machine learning technique that combines several decision trees to produce conclusions. This algorithm can be applied for both grouping and forecasting, such as categorizing data or calculating variables from various models to predict results. In Random Forest, many decision trees are built, and the output from each tree is then combined to obtain the final prediction. Random Forest is known for its ability to overcome various machine learning problems, both classification and regression [13].

The algorithm utilizes many decision trees built randomly, where each tree is created independently. The differences between trees and their independent nature make Random Forest more resistant to overfitting. The dataset splitting process involves grouping data into several different subsets, each having a specific role in the data analysis flow. Before training the classification model with Random Forest, the main data is first divided into two main parts: training set to build the model and testing set to test its performance [14]. To build an optimal decision tree model, the following workflow and mathematical formulations need to be applied [15].

1. Entropy Formula

The entropy formula is shown in equation

$$Entropy\ (S)\ =\ \Sigma^{n}{}_{i=1}\ -pi\ *\ log_2\,pi\ *\ \ldots\ldots\ldots\ldots\ldots(1)$$

Where:
- S is defined as all research data to be processed
- A functions as determining variables to be evaluated for each data point in S
- n is needed to divide S into smaller sub-groups for more focused analysis
- Pi is useful for understanding data distribution by showing the relative contribution of each sub-group to the total dataset.

2. Determining Gain using equation

$$Gain\ (S,A)\ =\ Entropy\ (S)\ -\ \Sigma^{n}{}_{i=1}\ |S_i|/|S|\ *\ Entropy(Si)\ *\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Where:
- S defines the entire data population that is the subject of research
- A serves as a key parameter to be evaluated for each element in S
- n is important for classifying data based on variations in parameter A values
- |Si| is useful for measuring the prevalence of each category, while |S| provides proportional context by showing the total dataset size.

III. RESULT

The results of this research reveal that the application of the Random Forest algorithm can be used to assess the nutritional status of toddlers at Tanah Pasir Primary Health Center, North Aceh, enabling malnourished toddlers to receive necessary interventions. This algorithm classifies nutritional status based on parameters such as age, weight, and height. Data originating from Tanah Pasir Primary Health Center is processed into a new dataset for classification purposes. The output of this application is information related to toddler nutritional categories that have been determined through the classification process. This research is designed to create a system capable of categorizing the nutritional status of toddlers in the region by utilizing the Random Forest algorithm. With this approach, the system can automatically predict nutritional status based on body measurement data such as age, weight, and height.

A. Problem Identification

The main problem addressed is the difficulty in determining the nutritional status of toddlers quickly and accurately in the Primary Health Center environment. Many toddlers are still not handled in a timely manner due to limitations in manual nutritional monitoring systems. The problems identified in this research include the lack of automatic systems for nutritional status classification in the Primary Health Center environment. Nutritional status assessment is still conducted manually, which potentially causes delays in handling malnourished toddlers. An efficient and accurate classification method is needed to process toddler anthropometric data. With this problem background, the Random Forest algorithm is implemented because it is capable of managing complex data and providing good classification results.

B. Data Analysis

The purpose of using data in this research is to train and test the toddler nutritional status classification model with the Random Forest algorithm. Data was obtained from Tanah Pasir Primary Health Center, North Aceh, including variables such as age (months), weight (kg), height (cm), and nutritional status categories (W/A, W/H, H/A) based on WHO standards. This dataset serves as the main foundation in developing the classification system.

The dataset used in this study comprises a total of 2,612 toddler records collected from Tanah Pasir Primary Health Center, North Aceh. The data includes key anthropometric measurements: age (in months), body weight (in kilograms), and body height (in centimeters). The age range of toddlers in the dataset spans from 0 to 60 months, representing the critical golden period of child development.

Nutritional status categories were determined based on three anthropometric indices following WHO standards: weight-for-age (W/A), height-for-age (H/A), and weight-for-height (W/H). Each toddler was classified into one of four categories: Good Nutrition, Undernutrition, Overnutrition, or Severe Malnutrition, based on their anthropometric measurements and corresponding z-scores.

Data preprocessing was conducted using Microsoft Excel to ensure data quality and consistency. This process included data cleaning to remove incomplete records, particularly those missing critical values for age, weight, or height. Data format standardization was performed to ensure compatibility with the Random Forest algorithm. Records with missing values in key variables were excluded from the analysis to maintain data integrity. Following the cleaning process, data labeling was conducted to determine nutritional class labels based on anthropometric index calculations.

The cleaned dataset was then split into training and testing sets using an 80:20 ratio. Specifically, 2,090 records (80%) were allocated for training the Random Forest model, while 522 records (20%) were reserved for testing and validation. This split ratio was chosen to provide sufficient data for model training while maintaining an adequate testing set for reliable performance evaluation. The random splitting method was employed to ensure unbiased representation of all nutritional status categories in both training and testing sets.

Random Forest was selected for this classification task due to its advantages in handling datasets with many features, resistance to overfitting, and ability to provide high accuracy in classification data. Each decision tree in Random Forest is built using random subsets of data and random subsets of features (attributes). The collective voting process from all trees is used to produce the final prediction, which enhances the model's robustness and reliability.

For research purposes, a toddler nutritional status database officially obtained from Tanah Pasir Primary Health Center in North Aceh Regency is used as the primary data source. The following is an example of infant nutrition data.

TABLE 1.
INFANT NUTRITION DATA

No	Attribute	Description	Example
1	NIK	Population Identification Number	1108132611180001
2	NAME	Toddler Name	misyari altaf h
3	GENDER	Gender (F=Female, M=Male)	M
4	Birth Date	Birth date in Year, Month, Date format	26-11-2018
5	Birth Weight	Toddler's weight at birth	3.00
6	Birth Height	Toddler's height at birth	49.0
7	Parent Name	Toddler's parent name	suryati
8	District	Village or Sub-district of toddler's origin	Tanah pasir
9	Regency/City	Regency of toddler's residence	Aceh Utara
10	Province	Toddler's province	Aceh
11	W/A	Weight for Age	Normal Weight
12	H/A	Height for Age	Normal
13	W/H	Weight for Height	Normal

The following is a table of children's data with parameters of age, weight, height, and labels.

TABLE 2.
CHILDREN DATA WITH PARAMETERS

No	Name	Age	Weight	Height (cm)	Label
1	Nadia agustina	12	6.5	70	Undernutrition
2	Aiswa Nahla	14	7	72	Undernutrition

3	Arsyia Falisha	13	5.5	68	Severe Malnutrition
4	M.Sultan Alfatih	15	8	74	Normal
5	M.Alkhabir	16	8.5	75	Normal
6	Muhammad Rizki	17	10	76	Overnutrition
7	Fauzil Haqiqi	18	11.5	80	Obesity
8	Zainal Aulia	14	9	74	Normal
9	Suci Nazira	12	7.2	70	Undernutrition
10	Marzia Kamila	15	11	78	Obesity

The following is the weight parameter table:

TABLE 3.
WEIGHT PARAMETER TABLE

Weight	Label
5.5	Severe Malnutrition
6.5	Undernutrition
8	Normal
8.5	Normal
10	Obesity
11.5	Obesity

a) Weight ≤ 8.5 → [5.5, 6.5, 8.0, 8.5] → Labels: Severe Malnutrition, Undernutrition, Normal.
b) Weight > 8.5 → [10.0, 11.5] → Labels: Overnutrition, Obesity.

In the weight parameter table, each label has kg values. The threshold value for the weight parameter is 8.5. If weight is below 5.5, it results in severe malnutrition label; if the value is above 5.5 and below 6.5, it results in undernutrition label; if the value is above 8.0 and less than or equal to 8.5, it results in normal label; if the value is above 8.5 and below or equal to 10.0, it results in overnutrition label; and if the value is more than 10.0, it results in obesity label. Threshold: 8.5 kg.

The following is the height parameter table with cm values for each label:

TABLE 4.
HEIGHT PARAMETER TABLE

Height	Label
68	Severe Malnutrition
70	Undernutrition
72	Undernutrition
74	Normal
76	Overnutrition
80	Obesity

a) Height ≤ 73 → [68, 70, 72] → Undernutrition, Severe Malnutrition
b) Height > 73 → [74, 76, 80] → Normal, Overnutrition, Obesity

In the height parameter table, each label has cm values. The threshold value for the height parameter is 73. If the value is less than or equal to 68, it is categorized as severe malnutrition. For values greater than 68 but not exceeding 70, it falls into the undernutrition category. Values above 70 up to a maximum of 73 are classified as normal nutrition. Values between 73 to 80 fall into the overnutrition category, while values exceeding 80 are classified as obesity. The threshold is set at 73 cm.

The following is the age parameter table with age values for each label:

TABLE 5.
AGE PARAMETER TABLE

Age	Label
12	Severe Malnutrition
13	Severe Malnutrition
14	Undernutrition
15	Normal
17	Overnutrition
18	Obesity

a) Age ≤ 14.5 → [12, 13, 14] → Undernutrition, Severe Malnutrition
b) Age > 14.5 → [15, 17, 18] → Normal, Overnutrition, Obesity

In the age parameter table, each label has age values. The threshold value for the age parameter is 14.5. If the value is less than 14, it results in severe malnutrition; if the value is 14 and below 14.5, it results in undernutrition; if the value is 14.5 and above but below 17, it results in overnutrition; and if above 17, it results in obesity.

The following is a test table with Normal results:

TABLE 6.
TEST TABLE WITH NORMAL RESULTS

NO	NAME	Age (months)	Weight (kg)	Height (cm)	Label
1	Alfarizki	15	8.2	73	Normal

The following are the prediction results:

TABLE 7.
PREDICTION RESULTS

Prediction Results
(Weight): 8.2 ≤ 8.5 → **Normal**
(Height): 73 ≤ 73 → **Undernutrition**
(Age): 15 > 14.5 → **Normal**

a) The weight value is 8.2, so the prediction result for the weight parameter is Normal because 8.2 is between 8.0 and 8.5
b) The height value is 73, so the prediction result for the height parameter is Undernutrition because 73 is between 70 and 74
c) The age value is 15, so the prediction result for the age parameter is Normal because 15 is between 14.5 and 15.

The following are the majority voting results:

TABLE 8.
MAJORITY VOTING RESULTS

Majority Voting
Normal: 2 votes
Undernutrition: 1 vote
Normal

From the voting results, Normal Nutrition has 2 votes while Undernutrition has 1 vote. Therefore, in testing Alfarizki's data, the result is Normal Nutrition.

The evaluation process begins by displaying the confusion matrix from the system results, then continues with calculating precision, recall, F1-Score, and accuracy metrics. These parameters are used to measure the effectiveness of the Random Forest-based classification system. The detailed calculations are as follows.

a. Precision

$$Precision = 0.94612068965517 \; x \; 100\% = 95\%$$

Of all system predictions stating that toddlers fall into a certain nutritional class (for example "Normal Nutrition"), 95% of these predictions are correct. This means the system rarely mistakenly thinks toddlers who are not actually in that class belong to it.

b. Recall

$$Recall = 0.97772828507795 \; x \; 100\% = 98\%$$

A recall of 98% means: Of all toddlers who actually belong to a certain nutritional class, the system successfully identifies 98% of them. So, this system very rarely misses cases of problematic nutrition (low false negative).

c. F1-Score

$$F1 - Score = 0.96166484118291 \; x \; 100\% = 96\%$$

With an F1-Score reaching 96%, it can be seen that: The balance between Precision (95%) and Recall (98%) values is very good. This indicator is particularly relevant for cases where it is necessary to consider both false positives and false negatives in a balanced way, such as in determining toddler nutritional status which is a sensitive matter.

d. Accuracy

$$Accuracy = 0.92911877394636 \; x \; 100\% = 93\%$$

An accuracy of 93% means: Of all test data entered, 93% are correctly classified by the system. However, accuracy can be misleading if the data is unbalanced (for example, too much "Normal").

C. System Implementation

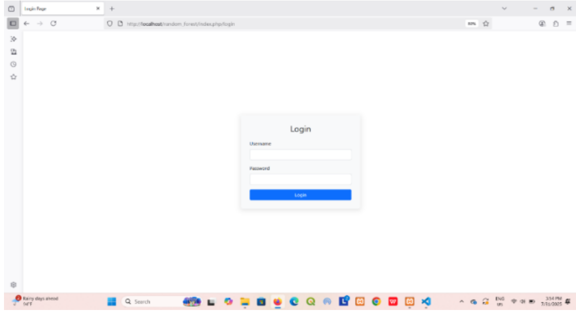

Figure 1. Login Page

The page above is the login page where users can enter their username and password to access the system. If successful, it will redirect to the dashboard page.

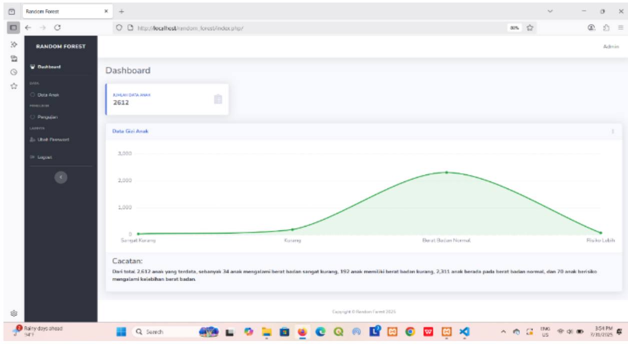

Figure 2. Dashboard Page

The image above shows the dashboard page. On this page, users can view several pieces of information in the system, namely the total count.

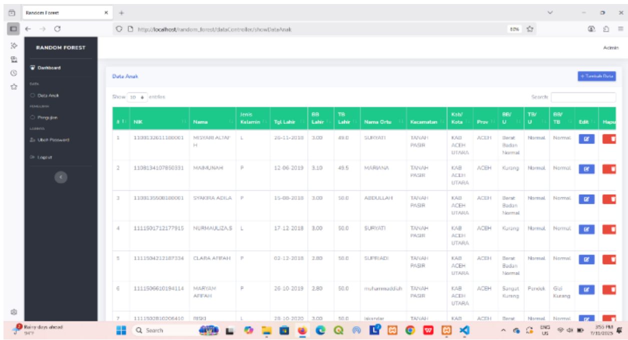

Figure 3. Children Data Page

The image above shows the children data page. This page is designed to manage children's data as a dataset for classification, perform preprocessing such as editing or manual validation before applying to the Random Forest algorithm, and serve as an important part of the child nutritional status detection or evaluation system.

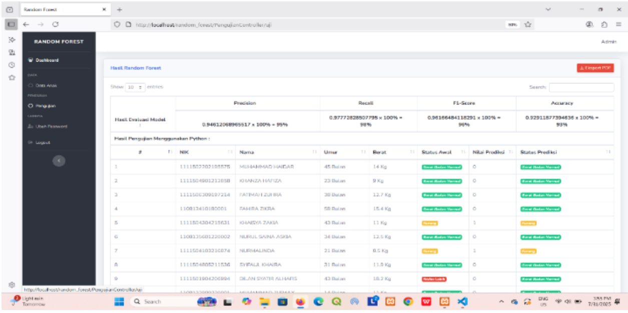

Figure 4. Testing Page

This display shows that the application has successfully implemented the Random Forest model for child nutritional status classification, clearly displays evaluation metrics, and provides individual prediction results, useful for decision-making in nutritional interventions.

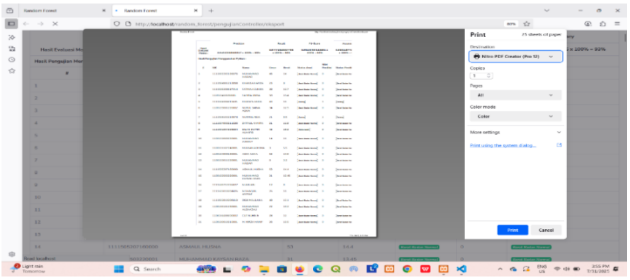

Figure 5. Export Page

The purpose of this feature is to document the results of child nutritional status classification testing. It is useful in reporting to health agencies, social services, or internal storage. The PDF format allows for neater and more formal storage and distribution.

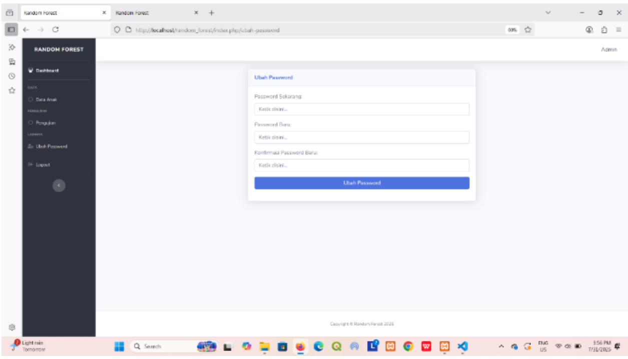

Figure 6. Change Password Page

This password change feature allows administrators to make changes independently as part of the system security mechanism. This capability is vital for protecting user accounts, especially when dealing with sensitive data processing such as toddler nutritional status records.

IV. Discussion

A. *Interpretation of Classification Performance*

The Random Forest model achieved a classification accuracy of 93%, which demonstrates strong performance in determining toddler nutritional status. In the context of public health screening, this accuracy level is considered highly acceptable and reliable for supporting clinical decision-making at primary health center settings. The 7% error rate, while present, is substantially lower than the potential for human error in manual assessment, particularly in high-volume screening scenarios where healthcare workers face time constraints and heavy workloads.

The precision value of 95% indicates that when the system predicts a toddler belongs to a specific nutritional category, it is correct 95% of the time. This high precision is particularly valuable in reducing false alarms, which could otherwise lead to unnecessary interventions or cause undue concern among parents and caregivers. Meanwhile, the recall of 98% demonstrates the model's exceptional ability to identify toddlers who actually belong to each nutritional status category, with only 2% of cases being missed. This high recall is especially critical in the context of malnutrition screening, where failing to identify at-risk toddlers (false negatives) could have serious health consequences, including delayed intervention and potential long-term developmental impacts.

The F1-score of 96% reflects an excellent balance between precision and recall, indicating that the model performs consistently well across all nutritional status categories without significantly favoring one metric over the other. This balanced performance is essential in healthcare applications where both false positives and false negatives carry distinct but equally important implications for patient care and resource allocation.

B. *Comparison with Manual Assessment Methods*

Traditional manual nutritional status assessment at Tanah Pasir Primary Health Center relies on healthcare workers manually measuring anthropometric parameters and cross-referencing them with WHO growth charts or reference tables. This process is time-consuming, prone to transcription errors, and dependent on the experience and workload of individual staff members. Based on informal observations during data collection, manual assessment typically requires 5-10 minutes per child, including measurement, chart consultation, and documentation.

In contrast, the Random Forest-based system can process and classify nutritional status in seconds once measurements are entered, providing immediate feedback to healthcare workers. This significant time reduction allows staff to screen more children within the same timeframe, particularly valuable during peak hours or community health campaigns. Moreover, the automated system provides consistent classification regardless of staff experience level or fatigue, reducing inter-observer variability that commonly affects manual assessments.

However, it is important to emphasize that this system is designed as a decision-support tool rather than a replacement for professional clinical judgment. Healthcare workers should use the system's predictions as a screening aid, with final decisions considering additional clinical factors such as medical history, recent illness, and family circumstances that the model does not capture.

C. Advantages and Limitations of Random Forest

The Random Forest algorithm proved particularly well-suited for this nutritional status classification task for several reasons. First, its ensemble approach, combining predictions from multiple decision trees, provides robust classification even when individual trees might misclassify certain cases. This robustness is valuable when dealing with borderline cases where anthropometric measurements fall near category boundaries.

Second, Random Forest's resistance to overfitting is crucial given the relatively modest dataset size of 2,612 records. While larger datasets would potentially improve performance further, the algorithm's ensemble nature and built-in regularization through random feature selection prevent the model from memorizing training data patterns that do not generalize well to new cases.

Third, the algorithm's ability to handle non-linear relationships between features naturally accommodates the complex interactions between age, weight, and height in determining nutritional status. Growth patterns in toddlers are inherently non-linear, and Random Forest captures these patterns without requiring explicit mathematical modeling of growth curves.

However, the study also acknowledges several limitations. First, the model's performance is tied to the quality and representativeness of the training data from Tanah Pasir Primary Health Center. Generalization to other regions with different demographic characteristics, socioeconomic conditions, or genetic backgrounds may require model retraining or recalibration. Second, the current implementation does not account for temporal changes, such as rapid weight loss due to illness or seasonal variations in food availability, which could affect classification accuracy in real-world deployment.

D. Clinical Implications of Classification Errors

In nutritional status classification, not all errors carry equal weight from a clinical perspective. False negatives—where the system fails to identify a malnourished child—are generally more concerning than false positives. A missed case of severe malnutrition could delay critical intervention, potentially resulting in irreversible developmental damage or, in extreme cases, mortality. The model's high recall of 98% specifically addresses this concern by minimizing the risk of missing at-risk children.

Conversely, false positives—where the system incorrectly flags a well-nourished child as malnourished—primarily result in unnecessary follow-up assessments. While this creates additional workload for healthcare staff and may cause temporary parental concern, these consequences are generally less severe than missing true malnutrition cases. The precision of 95% indicates that false positives occur in only about 5% of predicted malnutrition cases, maintaining a reasonable balance between sensitivity and specificity.

The system's threshold parameters (8.5 kg for weight, 73 cm for height, and 14.5 months for age) were derived from the training data distribution and align broadly with WHO growth standards adapted for the local population characteristics observed in the dataset. These thresholds should be periodically reviewed and potentially adjusted based on ongoing monitoring data and consultation with pediatric nutrition specialists.

E. Practical Implementation Considerations

Successful deployment of this classification system at Tanah Pasir Primary Health Center and potentially other health facilities requires addressing several practical considerations. First, healthcare workers need training not only in system operation but also in interpreting predictions and understanding the model's limitations. The system should clearly communicate prediction confidence levels, allowing staff to identify borderline cases requiring more careful evaluation.

Second, the system requires consistent and accurate anthropometric measurements as input. Measurement errors—such as inaccurate scales, improper height measurement technique, or incorrect age recording—will propagate through to classification errors. Standardized measurement protocols and regular equipment calibration are essential for maintaining system accuracy in practice.

Third, integration with existing health information systems and workflows is crucial for adoption. The web-based implementation with PDF export functionality facilitates documentation and reporting, but further integration with electronic medical records systems could enhance efficiency and enable longitudinal tracking of children's nutritional status over time.

F. Comparison with Alternative Machine Learning Methods

While this study focused on Random Forest, it is worth noting how this approach compares to alternative machine learning methods that could be applied to nutritional status classification. Decision Trees, the building blocks of Random Forest, offer excellent interpretability but are more prone to overfitting on training data. Support Vector Machines (SVM) can achieve high accuracy but require careful feature scaling and hyperparameter tuning, and their "black box" nature makes predictions harder to explain to healthcare workers.

Naive Bayes classifiers, while computationally efficient, assume feature independence—an assumption violated by the correlated nature of anthropometric measurements. Logistic Regression provides interpretable coefficients but may struggle with the non-linear relationships between age, weight, and height in determining nutritional status. Deep learning approaches like neural networks might achieve marginally better performance with larger datasets but require substantially more training data and computational resources while offering limited interpretability—a significant drawback in healthcare applications where explainability builds trust and adoption.

Random Forest strikes a favorable balance between accuracy, robustness, interpretability, and computational efficiency for this application. The feature importance metrics it provides can help healthcare workers understand which measurements most strongly influence classifications, supporting both trust in the system and educational opportunities about growth patterns.

G. Limitations of the Study

Several limitations should be acknowledged regarding this research. First, the dataset was collected from a single primary health center in Aceh Utara, potentially limiting generalizability to other regions with different population characteristics. Validation across multiple sites with diverse demographic profiles would strengthen confidence in the model's broader applicability.

Second, the study employed a single random train-test split rather than more robust validation approaches such as k-fold cross-validation or external validation on an independent dataset. While the 80:20 split provides a reasonable estimate of model performance, multiple random splits or stratified cross-validation would offer more reliable performance estimates and better assess model stability.

Third, the current implementation does not incorporate additional potentially relevant features such as feeding practices, maternal education, household income, or recent illness history, which are known to influence nutritional status but were not available in the dataset. Including such variables could potentially improve classification accuracy and provide more comprehensive risk assessment.

Fourth, the study does not examine model performance across different age groups or nutritional status categories individually. Some categories might be classified more accurately than others, and model performance might vary for different age ranges within the 0-60 month spectrum. Future work should investigate these performance variations to identify areas needing improvement.

H. Recommendations for Future Research

Based on the findings and limitations of this study, several directions for future research emerge. First, expanding the dataset to include multiple health centers across diverse geographic regions would enable assessment of model generalizability and identification of region-specific patterns requiring model adaptation.

Second, implementing longitudinal tracking where individual children's nutritional status is monitored over multiple visits would enable prediction of nutritional trajectories and identification of children at risk of declining status, facilitating proactive rather than reactive interventions.

Third, incorporating additional socioeconomic and clinical variables into the model could enhance prediction accuracy and provide more comprehensive risk assessment. Feature importance analysis could then identify the most critical factors influencing nutritional status in the local context, informing targeted intervention strategies.

Fourth, comparative studies rigorously evaluating Random Forest against other machine learning algorithms using the same dataset with proper cross-validation would provide empirical evidence for method selection rather than relying on theoretical advantages.

Finally, prospective validation studies where the system is deployed in real clinical settings and its predictions are compared with expert clinician assessments would provide the most convincing evidence of practical utility and identify operational challenges requiring attention before widespread adoption.

I. Broader Implications for Public Health

Beyond the immediate application at Tanah Pasir Primary Health Center, this research demonstrates the potential for machine learning approaches to strengthen nutritional surveillance and screening systems in resource-constrained settings. Automated classification systems can extend the effective reach of limited healthcare personnel, enabling more systematic and frequent monitoring of child nutritional status across populations.

The web-based implementation makes the system accessible from various devices and locations, potentially supporting community health workers conducting field screenings or enabling parents to input measurements for preliminary assessment before clinic visits. Such expanded access could improve early detection of nutritional problems and increase healthcare-seeking behavior among families concerned about their children's growth.

Furthermore, aggregated data from automated classification systems can provide valuable epidemiological insights for health planning and resource allocation. Identifying geographic or demographic patterns in malnutrition prevalence can guide targeted intervention programs and enable more efficient distribution of nutritional supplementation resources.

However, it is crucial to acknowledge that technological solutions like this classification system address only one component of the complex, multifaceted challenge of child malnutrition. Sustainable improvements in child nutritional status require comprehensive approaches addressing underlying determinants including food security, clean water access, sanitation, maternal education, and poverty reduction. The classification system should be viewed as a tool supporting, but not replacing, these broader public health interventions.

V. CONCLUSION

The research and implementation of a system for classifying nutritional status of toddlers using the Random Forest algorithm yielded several important findings. The developed system effectively utilized the Random Forest approach to determine children's nutritional categories based on age, weight, and height criteria, following three assessment standards: weight-for-age (BB/U), weight-for-height (BB/TB), and height-for-age (TB/U). Random Forest demonstrated its capability in performing classification effectively, with system testing results showing evaluation metrics of 93% accuracy, 95% precision, 98% recall, and 96% F1-score. These results prove that Random Forest is reliable in detecting toddlers with poor or severely poor nutritional status. The system facilitates primary health centers in detecting and managing toddler nutritional status data quickly and efficiently. To enhance efficiency and accuracy, the system has transitioned from manual methods to more structured and documented automatic processes. The objective of developing this web-based system is to create a better data storage solution with superior features such as PDF export, centralized data management, and security mechanisms through updatable password settings.

REFERENCES

[1] E. N. Candra, I. Cholissodin, and R. C. Wihandika, "Klasifikasi Status Gizi Balita menggunakan Metode Optimasi Random Forest dengan Algoritme Genetika (Studi Kasus : Puskesmas Cakru)," vol. 6, no. 5, pp. 2188–2197, 2022.

[2] M. Muhammad, A. Mahmudi, and K. Auliasari, "Perbandingan Metode K-Means Dan K-Medoids Untuk Klasifikasi Status Gizi Anak," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 4, pp. 2122–2129, 2023, doi: 10.36040/jati.v7i4.7403.

[3] M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto, and D. Suhendro, "Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi," *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 2, no. 1, pp. 30–38, 2021, doi: 10.37148/bios.v2i1.19.

[4] C. Zai, "Implementasi Data Mining Sebagai Pengolahan Data," 2022.

[5] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 1, Apr. 2021, doi: 10.30813/jbase.v4i1.2729.

[6] Fauziah, Dedy Hartama, and Irfan Sudahri Damanik, "Analisa Kepuasan Pelanggan Menggunakan Klasifikasi Data Mining," *Jurnal Penerapan Kecerdasan Buatan*, 2020.

[7] D. M. Musa *et al.*, "Penerapan Data Mining Untuk Klasifikasi Data Penjualan Pakan Ternak Terlaris Dengan Algoritma C4.5," *Jurnal Teknologi Informatika dan Komputer*, vol. 10, no. 1, pp. 168–182, Mar. 2024, doi: 10.37012/jtik.v10i1.1985.

[8] B. W. K. Nurdin, "Implementasi Data Mining Untuk Mengklasifikasi Data Nasabah Pt. Adira Finance Aceh Tengah Menggunakan Algoritma C4.5," *Jurnal Sistem Informasi Kaputama (JSIK)*, vol. 1, no. 1, 2017.

[9] R. A. Wati, H. Irsyad, M. Ezar, and A. Rivan, "Klasifikasi Pneumonia Menggunakan Metode Support Vector Machine," 2020.

[10] P. B. N. Setio, D. R. S. Saputro, and B. Winarno, "Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prosiding Seminar Nasional Matematika*, vol. 3, pp. 64–71, 2020, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[11] H. I. Islam, M. Khandava Mulyadien, U. Enri, U. Singaperbangsa, and K. Abstract, "Penerapan Algoritma C4.5 dalam Klasifikasi Status Gizi Balita," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 10, pp. 116–125, 2022, doi: 10.5281/zenodo.6791722.

[12] T. Wang, "Improved random forest classification model combined with C5.0 algorithm for vegetation feature analysis in non-agricultural environments," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-60066-x.

[13] B. Prasojo and E. Haryatmi, "Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 79–89, Sep. 2021, doi: 10.25077/teknosi.v7i2.2021.79-89.

[14] N. G. Ramadhan, F. D. Adhinata, A. J. T. Segara, and D. P. Rakhmadani, "Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 251, Apr. 2022, doi: 10.30865/jurikom.v9i2.3979.

[15] A. P. Ruise, A. S. Mashuri, M. Sulaiman, and F. Rahman, "Studi Komparasi Metode Svm, Logistic Regresion Dan Random Forest Clasifier Untuk Mengklasifikasi Fake News di Twitter," *J I M P - Jurnal Informatika Merdeka Pasuruan*, vol. 7, no. 2, p. 64, Sep. 2023, doi: 10.51213/jimp.v7i2.472.