INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 17 No. 1 (2025) 463-475
https://publisher.uthm.edu.my/ojs/index.php/ijie

Performance Evaluation of Different Classification Algorithms Applied for Identifying Maternal Nutritional Status by Anthropometric Measurements

Diva Kurnianingtyas[1], Nathan Daud[1], Agus Wahyu Widodo[1]*, Tutut Herawan[2]

[1] *Department of Informatics Engineering, Faculty of Computer Science, Universitas Brawijaya, Malang, 65154, INDONESIA*
[2] *Department of Information System, Faculty of Computer Science and Information Technology, University of Malaya, 50603, MALAYSIA*

*Corresponding Author: a_wahyu_w@ub.ac.id
DOI: https://doi.org/10.30880/ijie.2025.17.01.037

Article Info

Received: 5 September 2024
Accepted: 10 April 2025
Available online: 30 April 2025

Keywords

Artificial intelligence, machine learning, optimization, stunting, pregnancy, pregnant women, nutrition

Abstract

Pregnancy significantly influences infant quality and development. Maternal monitoring, indicated by body mass index (BMI) and mid-upper arm circumference (MUAC) measurements, reflects a country's socioeconomic development. Improper measurements heighten the risk of chronic energy deficiency (CED) in pregnant women and low birth weight (LBW) in infants. This study leverages artificial intelligence (AI) to enhance the detection process. Specifically, it evaluates the prediction performance of various classification methods: Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM). Using interviews in Jombang District, Indonesia, these methods were expected to identify maternal nutritional status. The model design was divided into two stages: MUAC estimation generated binary classes, and BMI estimation generated multiple classes. The evaluation of these methods included various performance metrics: Accuracy (Acc), G-means, Sensitivity (Sens), Specificity (Spec), Positive Predictive Value (PPV), and Negative Predictive Value (NPV). Based on the results, all methods are proposed for both classifications, except KNN on multiple classification. KNN achieved significant scores in all matrices with $p<0.01$. KNN's performance is impacted by data imbalance. The study revealed a strong correlation (0.92 coefficient) between BMI and MUAC variables. The application of ML algorithms in detecting maternal nutritional status can significantly enhance the effectiveness and efficiency of health facilities, especially in areas with inadequate resources and medical personnel. However, exploring diverse ML algorithms is recommended to find optimal approaches for more varied data and to contribute solutions for sustainable development in the country.

1. Introduction

The nutritional status of pregnant women is a measure of success in providing nutrition for mothers. Balanced nutrition in pregnancy is crucial because the maternal nutritional status significantly affects the baby's condition in the womb with the maternal food intake passing to the fetus through the umbilical cord. In Indonesia, maternal health challenges remain pressing, with 20.7% of pregnant women experiencing chronic energy deficiency (CED) and 48.9% suffering from anemia, which underscores the need for targeted nutritional interventions to improve outcomes for both mothers and babies. Adequate food intake is required for the fulfilment of the fetus's nutritional needs, ensuring optimal fetal growth and development [1]. Improper maternal nutritional status may risk maternally diagnosed CED and nutritional anemia [2], as well as giving birth to a low birth weight baby (LBW).

LBW poses risks to both mother and baby and is associated with an increased risk of infant morbidity and mortality, delayed growth and cognitive development, and the possibility of suffering from chronic diseases later in life [3]. The risk of having an LBW baby increases due to nutritional problems or CED that occur in eligible women (EW) [4]; these problems can inhibit fetal growth during pregnancy.

The detection of maternal nutritional status can be performed using anthropometric measurements taken by calculating body mass index (BMI) and mid-upper arm circumference (MUAC) [5]. BMI calculation detects four maternal nutritional statuses: obese, overweight, normal, and underweight. Conversely, MUAC calculation can determine the chances of a pregnant woman being at risk or not of developing CED [6].These two measurements are used as essential indicators to avoid LBW.

The risk of LBW can be minimized with technology and the development of machine learning (ML) [7] has been helpful for identification and prediction in the healthcare industry. ML is effective in diagnosing, predicting patient outcomes, recognizing patterns in medical data, and improving the effectiveness of treatment [8]. Some researchers have also implemented the technology to predict BMI status. Researchers [9] believe that BMI status combined with voice features can help treat patients with prognosis of diseases such as stroke, diabetes, and cardiovascular diseases through the use of logistic regression (LR), bagging, and random forest (RF) algorithms. The results show that LR is better than both ensemble methods using imbalanced data. However, this contrasts research [10], evidencing that the tree-based algorithm provides the best accuracy results for BMI status prediction using data from patient CT scans. Research [11] has also determined BMI status based on psychological factors using eight algorithms, including K-nearest neighbor (KNN), classification and regression tree (CART), support vector machine (SVM), multi-layer perceptron (MLP), Ada boosting with decision tree (AB), gradient boosting (GB), random forest (RF), and extra tree (ET). The results obtained show that all algorithms have an accuracy above 80%. Regarding nutritional status, researchers [12], [13] have mostly examined the relationship between MUAC and BMI, rather than predicting the nutritional status of patient health. However, studies to identify the maternal nutritional status are still needed. This study is important to minimize the risk of stunting because LBW can potentially disrupt human resource capacity and is associated with health levels, including child mortality [14], a concern of the World Health Organization (WHO).

Based on the previous explanation, a study is needed to identify the maternal nutritional status with the best performance algorithm. The main objective of this study was to explore alternative methods for detecting maternal nutritional status. To solve this problem, classification models such as Decision Tree (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM), which have different numbers of classification categories, are applied to calculate BMI and MUAC. This experiment will be able to maximize the findings from the performance evaluation results of each model.

This study is divided into sections. Section 2 describes the steps to identify maternal nutritional status while the following section (Section 3) shows the results obtained with several classification methods. Section 4 presents the results of the comparison of the methods used to detect maternal nutritional status, and, finally, explains the conclusions and limitations for future research

2. Methods

The methods section of the study outlines the processes and techniques used to evaluate the performance of different classification algorithms in detecting maternal nutritional status. The study involves data collection from interviews with pregnant women in Jombang District, Indonesia, resulting in a dataset of 100 samples. This dataset includes variables such as age, parity, and anthropometric measurements like MUAC and BMI. The classification algorithms employed are DT, KNN, LR, NB, RF, and SVM. Each algorithm was optimized through hyperparameter tuning to enhance predictive accuracy. The study implemented tenfold cross-validation to assess the models' performance using metrics such as accuracy, sensitivity, specificity, positive and negative predictive values, and geometric mean.

2.1 Study Sample and Variables

Sample data were sourced from direct interviews with pregnant women in Jombang District, Indonesia. The information collected comprised 100 data. Categorization of maternal nutritional status derived from studies [15], [16] and the dataset characteristics in this study are shown in Table 1.

Table 1 *Characteristics of data in this study*

Variable	n=100	
Age	Mean±Std	27.43±6.18
	Median	26.0
	Range (Min-Max)	17.0-40.0
Parity	Mean±Std	1.86±1.37
	Median	2.0
	Range (Min-Max)	0.0-4.0
Preg_Interval	Mean±Std	16.92±6.98
	Median	16.5
	Range (Min-Max)	1.0-28.0
MUAC	Mean±Std	24.88±7.04
	Median	24.0
	Range (Min-Max)	18.0-37.0
Height	Mean±Std	153.02±8.03
	Median	153.0
	Range (Min-Max)	140.0-166.0
Weight	Mean±Std	56.11±4.52
	Median	56.0
	Range (Min-Max)	45.0-70.0
BMI	Mean±Std	24.45±3.12
	Median	24,415
	Range (Min-Max)	16.33-33.08
Risk Status	No (≥23.5 cm)	52
	Yes (<23.5 cm)	48
BMI Status	Underweight (<18.5 kg/m^2)	27
	Normal (18.5-24.9 kg/m^2)	25
	Overweight (25-29.9 kg/m^2)	24
	Obese (>30 kg/m^2)	24

2.2 Classification Algorithms

Researchers have implemented several methods to solve the problems, such as DT, KNN, LR, NB, RF, and SVM. These methods are expected to help the process of detecting maternal nutritional status. DT can classify and predict by constructing a tree from training data in the form of cases recorded in the database. Each case consists of the values of the attributes of the class. Each attribute is either discrete or continuous data but cannot be empty [17]. KNN requires distance calculation between training data and new objects for classification. The k value limits the number of nearest neighbors in the distance calculation as a consideration in the classification process. LR provides the relationship between the independent and dependent variables when the dependent variable is categorical. The regression process continues if there are still insignificant independent variables [18]. NB uses probability and statistical methods so that there is no influence between attributes. High accuracy and speed are advantages in applying to databases with extensive data. RF applies bootstrap aggregating or bagging and random feature selection methods. Algorithms with predictive decision techniques with random predictors that combine results with a majority vote are applied [19]. SVM performs classification, regression, and outline detection with a hyperplane affording the advantages of effective processing in high dimensional space and efficient use of memory [20]. The classification model was constructed and executed using the Python library package

2.3 Method Comparison

An evaluation process to select the best model is required in AI-related projects. The evaluation components that need to be considered in this study are sensitivity (Sens), specificity (Spec), geometric mean (G-means), positive predictive value (PPV), negative predictive value (NPV), and accuracy (Acc) [21]. Sens is used to calculate the test accuracy of positive claims on the maternal nutritional status. Next, Spec is the ability to estimate the accuracy of the test on positive responses to the maternal nutritional status. G-means are applied for imbalance class measurement by combining sensitivity and specificity values. Furthermore, PPV is the ability to predict how likely a positive nutritional status with test results is true in the population. NPV is the degree of prediction to assess how likely a negative test outcome with the test index is true in the population. Acc is used to evaluate the accuracy of diagnostic of maternal nutritional status. These metrics are crucial in clinical decision-making; for example, high sensitivity ensures that most at-risk pregnant women are correctly identified for early intervention, while high specificity minimizes unnecessary treatments for those not at risk. In addition, the ML algorithm is compared with the area under the curve (AUC) calculation approach. AUC is used to measure the comparison of the performance of model prediction outcomes with a range of 0.0 to 1.0. When the AUC score is close to 1.0, the algorithm is capable of performing a satisfying classification. Evaluation calculation used the Python library package.

The study will conduct two experiments on the classification model using the same datasets. Datasets in the first experiment have two categories, which are binary classification. In comparison, the other experiment has four labels (multi-class classification). The illustration is shown in Fig. 1.

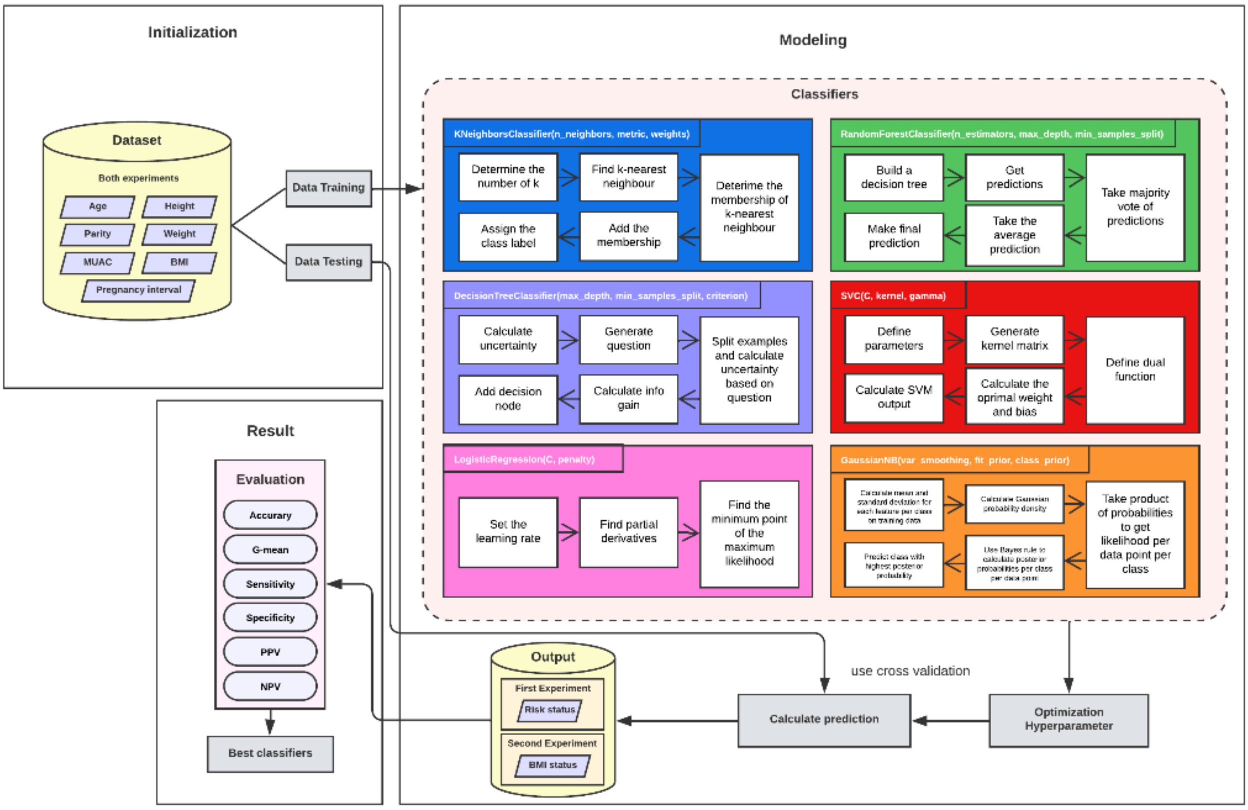

Fig. 1 *Nutritional status detection design*

3. Results

The results section presents the findings from the application and evaluation of the machine learning algorithms. The exploratory analysis highlights significant correlations between variables influencing maternal nutritional status, notably a strong positive correlation between BMI and MUAC (r = 0.92). In binary classification (identifying risk status), all algorithms performed well, with no statistically significant differences among them, indicating their suitability for detecting chronic energy deficiency using MUAC measurements. However, in multiclass classification (BMI categorization), KNN showed significant performance issues compared to other methods, attributed to its vulnerability to data imbalance and high dimensionality. The Receiver Operating Characteristic (ROC) curves supported these findings, with all algorithms except KNN achieving high AUC values in multiclass classification. Overall, the study suggests that while most algorithms can effectively classify maternal nutritional status, KNN is less reliable in multiclass scenarios.

3.1 Exploratory Analysis

ML approach is a method for processing large amounts of historical data and identifying patterns that characterize the data. ML methods can also help make more accurate results predictions based on a given set of input data. In its process, ML involves the process of recognizing the correlation between variables. This stage is used as data preparation so that the collected data can determine whether the data is appropriate for the problem to be solved. Furthermore, the results are used to clean data on inconsistent and incomplete data. Based on this explanation, inaccurate information on maternal nutritional problems results in inappropriate menu recommendations. Inappropriate nutritional intake affects the condition of the baby in the womb. The characteristics of variables that influence the maternal nutritional status are presented in Fig. 2. There is a significant difference in all predictor variables that impact the response variable represented by *risk_status* and *BMI*.

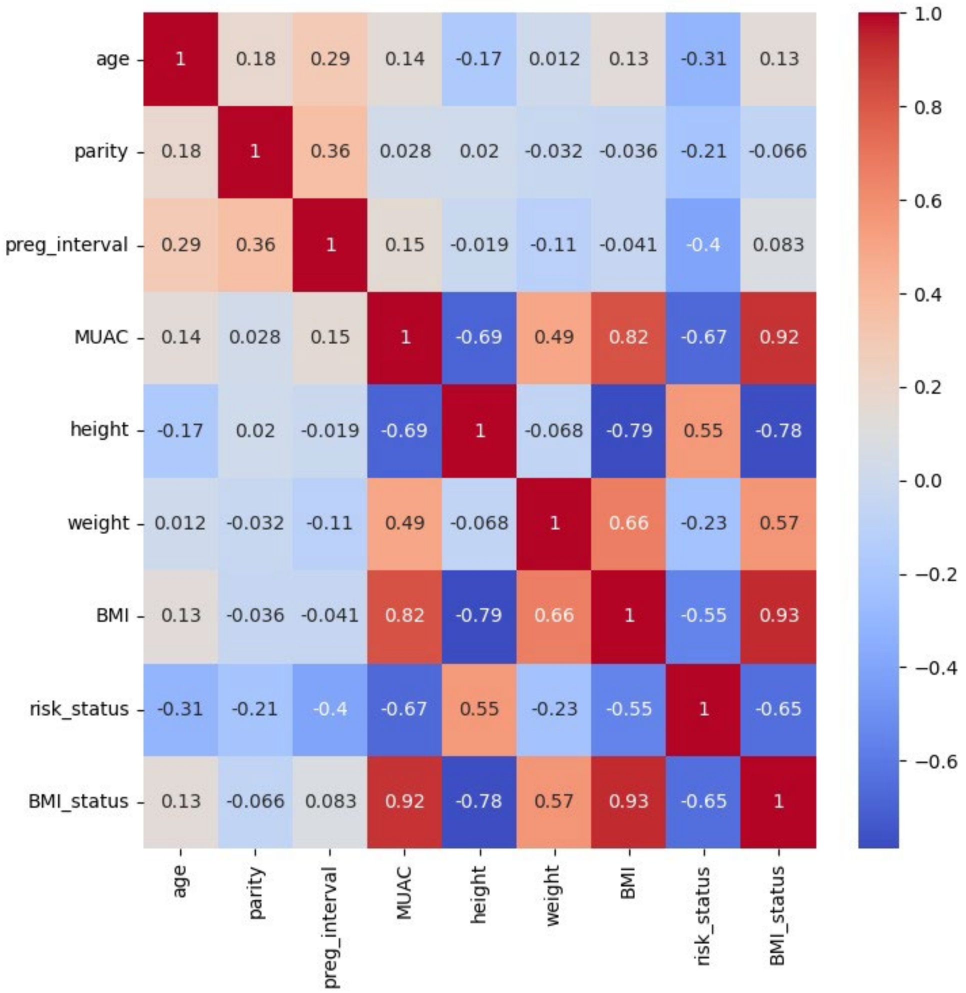

Fig. 2 *Variable characteristics in this study*

The *risk_status* indicator focuses on the determination of whether a pregnant woman is at risk or not (binary classification) of getting CED, which *MUAC* typically measures. Additionally, *risk_status* had the highest negative correlation with *MUAC*, with a coefficient value of -0.67. The *risk_status* variable is declared risky when the *MUAC* score is below the requirement in Table 1, and, conversely, it is stated to be negatively correlated. The correlation was proven in a study [22], which showed that the *MUAC* value decides the maternal risk status.

BMI status is applied to identify maternal nutritional status in the second experimental or multi-class classification label experiment. The strong positive correlation between BMI and MUAC (r = 0.92) aligns with prior studies that highlight MUAC as a reliable proxy for BMI in assessing maternal nutritional status, particularly in resource-limited settings where BMI calculations may be less practical. Based on the findings, *BMI_status* has a positive correlation with *BMI, MUAC, weight, age*, and *preg_interval* using coefficient values of 0.93, 0.92, 0.57, 0.13, and 0.083. *BMI_status* categorization is based on the requirements given in Table 1. Studies have also evidenced the relationship between *BMI_status* and *MUAC* [12]. In addition, the correlation concluded that BMI rises with age. This result is influenced by an increase in body weight at first pregnancy. Also, short pregnancy spacing has a significant obesity risk. Uncontrolled obesity raises the onset of chronic diseases in old age [23]. This statement is supported by *BMI_status*, which has a negative correlation with *parity*, *risk_status*, and *height* with coefficient values of -0.066, -0.65, and -0.78, respectively.

3.2 Trained Classification Methods

The learning process in ML requires training parameters to control the quality of the outcome model. Model hyperparameters are adjusted manually and utilized in the estimation process to facilitate the prediction of model parameters. While hyperparameter settings generally generate useful models by default, undertaking hyperparameter tuning enhances the predictiveness of the model. Hyperparameter tuning, a crucial process of finding optimal model settings through systematic parameter adjustments, enhances the algorithms' ability to accurately detect maternal nutritional status by minimizing prediction errors and improving overall model performance.

Hyperparameter tuning is conducted on DT, KNN, LR, NB, RF, and SVM algorithms. This study performs hyperparameter tuning on binary and multiclass classification on the six methods shown in Fig. 4. The tuning results are based on the variation of hyperparameter values corresponding to the classifier. The decision is based on a consistent Mean Absolute Percentage Error (MAPE) value. DT, using max_depth, found the position of max_depth=5 in binary classification and max_depth=3 in multiclass classification. The testing on KNN resulted in the values k=1 and k=5 in binary and multiclass classification. Furthermore, LR is based on the variable C and obtaining the value of C=0.1 in binary and multiclass classification. NB locates the var_smoothing position based on the minimum MAPE value equal to 1e-1 and 1e-2 in binary and multiclass classification. Also, RF acquires the value of n_estimators=100 in binary and multiclass. SVM applies the value of C to find the optimal position. In binary classification, the minimum MAPE value occurs at position C=1000, while for multiclass classification, it is at C=1000. The hyperparameter tuning results will be used to compare the performance of the six classifiers.

3.3 Method of Comparison

The performance of different classification models for the identification of maternal nutritional status was performed with tenfold cross-validation (CV) to obtain the Sens, Spec, PPV, NPV, G-means, and Acc values, as shown in Fig. 5. Performance comparison with tenfold CV generated diverse results and thus can be segmented into a box. Based on the results in binary classification, the evaluation outcomes provide scores that are not significantly different because the maximum value of Sens, Spec, PPV, NPV, G-means, and Acc, which is 1.0. In addition, there are no outliers in any of the classification clusters. Meanwhile, in multiclass classification, KNN has a significantly different value compared to other classification algorithms. The insufficient performance of KNN in multiclass classification is limitation of KNN. The capabilities of KNN weaken when the dimensionality increases and the proximity approaches the average distance between points [24]. In addition, results regarding the comparison of operating characteristics (OC) between paired algorithms with significant differences are reported in Table 2 and Table 3.

The receiver operating characteristic (ROC) curve provides a way to examine classification algorithms for identification. In the medical sector, ROC curves are widely utilized to visualize diagnostic accurateness and determine optimal cut-off values [25]. Based on Fig. 3, all the methods achieved AUC>0.9 in binary classification. The detection procedure is considered outstanding because the value acquired is closer to a 1.0 value. In the multiclass classification, KNN only gained an AUC value of 0.80 compared with other methods that achieved AUC>0.9. As explained earlier, KNN is more capable of handling lower data dimensions.

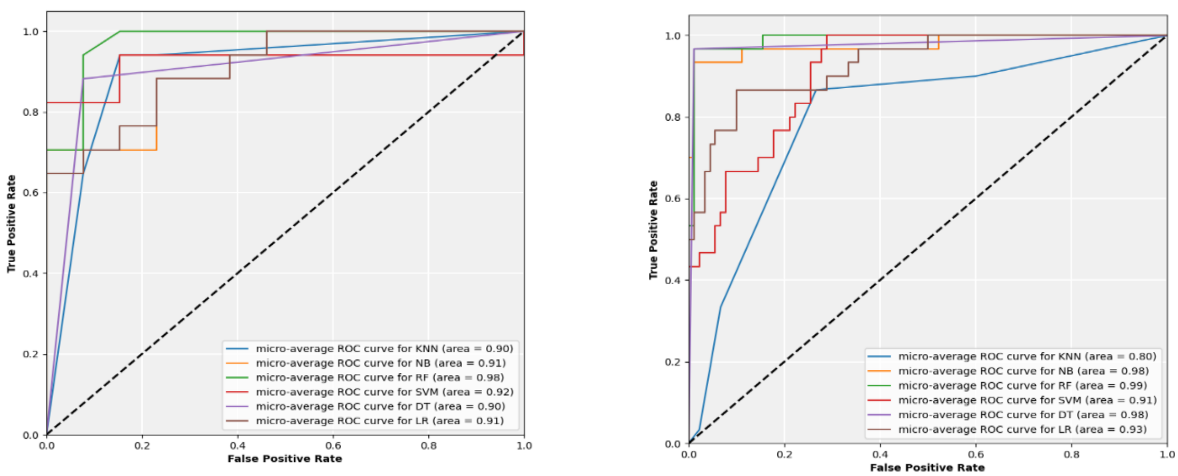

Fig. 3 *ROC curve for the classifiers*

Int. Journal of Integrated Engineering Vol. 17 No. 1 (2025) p. 463-475

4. Discussion

Six ML methods, consisting of DT, KNN, LR, NB, RF, and SVM, have been implemented and tested in the study of maternal nutritional status detection. Through ten-fold CV (see Fig. 5), ROC curve (see Fig. 3), and pairwise comparison tests (see Table 2 and Table 3, several characteristics of the classifiers were obtained that can be discussed.

Table 2 *Pairwise comparison test among different classification methods on binary classification*

	Acc	*G-mean*	*Sens*	*Spec*	*PPV*	*NPV*
KNN-NB	p=.87	p=.85	p=.77	p=.96	p=.89	p=.70
KNN-RF	p=.20	p=.23	p=.20	p=.84	p=.69	p=.08
KNN-SVM	p=.52	p=.52	p=.31	p=.96	p=.92	p=.23
KNN-DT	p=.74	p=.73	p=.45	p=.83	p=1.00	p=.40
KNN-LR	p=.74	p=.75	p=.77	p=.89	p=.84	p=.64
NB-RF	p=.21	p=.26	p=.29	p=.78	p=.55	p=.15
NB-SVM	p=.60	p=.62	p=.44	p=1.00	p=.96	p=.39
NB-DT	p=.85	p=.87	p=.63	p=.77	p=.88	p=.65
NB-LR	p=.86	p=.90	p=1.00	p=.83	p=.71	p=.94
RF-SVM	p=.46	p=.49	p=.75	p=.77	p=.58	p=.57
RF-DT	p=.24	p=.31	p=.53	p=1.00	p=.64	p=.26
RF-LR	p=.26	p=.30	p=.29	p=.96	p=.83	p=.16
SVM-DT	p=.70	p=.73	p=.75	p=.75	p=.92	p=.63
SVM-LR	p=.71	p=.71	p=.44	p=.83	p=.74	p=.42
DT-LR	p=1.00	p=.98	p=.63	p=.95	p=.82	p=.70

Table 3 *Pairwise comparison test among different classification methods in multiclass classification*

	Acc	*G-mean*	*Sens*	*Spec*	*PPV*	*NPV*
KNN-NB	p<.01	p<.01	p<.01	p<.01	p<.01	p<.01
KNN-RF	p<.01	p<.01	p<.01	p<.01	p<.01	p<.01
KNN-SVM	p<.01	p<.01	p<.01	p<.01	p<.01	p<.01
KNN-DT	p<.01	p<.01	p<.01	p<.01	p<.01	p<.01
KNN-LR	p<.01	p<.01	p<.01	p<.01	p<.01	p<.01
NB-RF	p=.57	p=.86	p=.78	p=.90	p=.81	p=.81
NB-SVM	p=.78	p=.80	p=.78	p=.93	p=.96	p=.96
NB-DT	p=.79	p=.91	p=1.00	p=.66	p=.89	p=.89
NB-LR	p=.26	p=.27	p=.26	p=.32	p=.32	p=.32
RF-SVM	p=.34	p=.94	p=.53	p=.81	p=.80	p=.80
RF-DT	p=.39	p=.76	p=.75	p=.74	p=.69	p=.69
RF-LR	p=.09	p=.32	p=.15	p=.36	p=.20	p=.20
SVM-DT	p=1.00	p=.69	p=.75	p=.55	p=.83	p=.83
SVM-LR	p=.33	p=.34	p=.33	p=.22	p=.23	p=.23
DT-LR	p=.36	p=.20	p=.23	p=.57	p=.36	p=.36

When comparing classification models on binary classification, based on Table 2, the Acc, G-means, Sens, Spec, PPV, and NPV values obtained $p>0.05$. These values explain that the methods' performance is different from each other, but all methods can detect the maternal nutritional status appropriately. This statement has been proven in the ten-fold CV test therein the maximum scores on the Acc, G-mean, Sens, Spec, PPV, and NPV values are 1.0, with various distributions of target data earned by each classifier.

KNN exhibited poor performance in multiclass classification, with a maximum accuracy of 0.80 and a significant p-value ($p<0.01$). The boxplot representation confirms its limitations, showing suboptimal test scores compared to other classifiers. While previous studies [26], [27] reported KNN's success, its sensitivity to data imbalance negatively impacted prediction accuracy in this study. The imbalanced distribution of BMI status categories led KNN to favor majority class predictions, preventing it from achieving an accuracy of 1.0. Similar shortcomings were observed in [28], where KNN struggled with high-dimensional, imbalanced data due to its reliance on distance metrics. Future research should explore resampling strategies like SMOTE to mitigate these effects. Furthermore, this study revealed an association between BMI status and MUAC values. The results can be recommended to medical personnel to detect the maternal nutritional status using BMI calculations or measuring MUAC. However, MUAC is usually preferred because it is easier to practice than BMI [29].

Based on the observed results, it is hardly possible to justify a small dataset in the context of a country's problem. The relatively small sample size (N = 100) in this study may limit the generalizability of the findings, and future research involving larger and more diverse datasets across different regions in Indonesia is recommended to enhance the robustness and applicability of the results. In addition, other criteria need to be considered to enrich the classification and detection process. More research is needed that explores more ML algorithms to find an effective and efficient algorithm for detecting maternal nutritional status with larger and more varied data. This solution can contribute to the sustainable development of the country [30]. Future research should explore deep learning architectures, ensemble methods, and neural networks while incorporating additional variables such as socioeconomic factors, dietary patterns, and longitudinal health data to develop more robust predictive models for maternal nutritional status across diverse Indonesian populations.

Several findings can be drawn from this study related to the performance of the ML algorithms being compared. Based on the results, in binary classification, there is no significant difference in observations because the values of Acc, Spec, Sens, NPV, and PPV reached $p>0.01$. The p-value is helpful to validate the hypothesis against the observed data. The study's initial hypothesis states the possibility that DT, KNN, LR, NB, RF, and SVM methods can identify the maternal status successfully. The hypothesis was accepted, as evidenced by a p-value of >0.01. Therefore, all ML algorithms can recognize nutritional status successfully, especially to identify the maternal risk status. Nevertheless, the KNN method has a weak performance in multiclass classification in OC term, which has a p-value <0.01 compared to other methods. Vulnerability to data imbalance is the finding of this study as a shortcoming of KNN. The KKN capability weakens as the dimension rises, and the proximity closes to the average distance between points.

All ML methods are proposed for binary and multiclass classification, except for KNN, which can only be applied to binary classification. The implementation of ML methods can enhance effectiveness and efficiency in health facilities which are still inadequate and insufficient in the number of medical personnel. In addition, many ML algorithms need to be explored to find more effective and efficient algorithms for more varied data to assist in the detection of maternal nutritional status in supporting the sustainable development of the country. The successful implementation of these ML algorithms can inform evidence-based policy decisions for maternal health programs, particularly in resource-limited areas of Indonesia, by enabling targeted resource allocation, improving early intervention strategies, and strengthening the capacity of local healthcare facilities to prevent maternal nutritional deficiencies. One way to increase accuracy is to discover the best hyperparameters of each method. The increasing amount of data that will continually occur does not allow for repeated searches for the best hyperparameters. This repetition will reduce the level of efficiency in identifying maternal nutrition. Therefore, it needs to be developed adaptively by implementing optimization approaches.

The study's focus on six classical machine learning algorithms (DT, KNN, LR, NB, RF, SVM) introduces limitations in algorithmic scope, particularly the exclusion of modern ensemble methods and deep learning architectures that could potentially improve classification performance for maternal nutritional status. While KNN demonstrated adequate performance in binary classification (AUC >0.9), its significant underperformance in multiclass scenarios (AUC=0.80) highlights sensitivity to class imbalance and high dimensionality - a limitation exacerbated by the dataset's uneven BMI category distribution (underweight:27, normal:25, overweight:24, obese:24). The constrained sample size (N=100) from a single Indonesian district limits generalizability across diverse populations with varying socioeconomic and nutritional profiles. Furthermore, the manual hyperparameter tuning approach (e.g., k=1 for KNN, max_depth=3 for DT) proves unsustainable for scaling to larger datasets, necessitating automated optimization strategies. The strong BMI-MUAC correlation (r=0.92) suggests clinical utility but overlooks critical micronutrient and socioeconomic determinants of maternal health. Future research should prioritize three key advancements: 1) Implementation of adaptive hyperparameter optimization using meta-learning approaches to maintain efficiency with expanding datasets, 2) Integration of synthetic minority oversampling (SMOTE) to address class imbalance in multiclass classification, and 3) Expansion of predictive features to include hematological markers and socioeconomic indicators alongside anthropometric measurements. Comparative studies across Indonesia's diverse regions using deep learning architectures could establish robust, population-specific classification models while maintaining MUAC's practical advantages in resource-constrained settings.

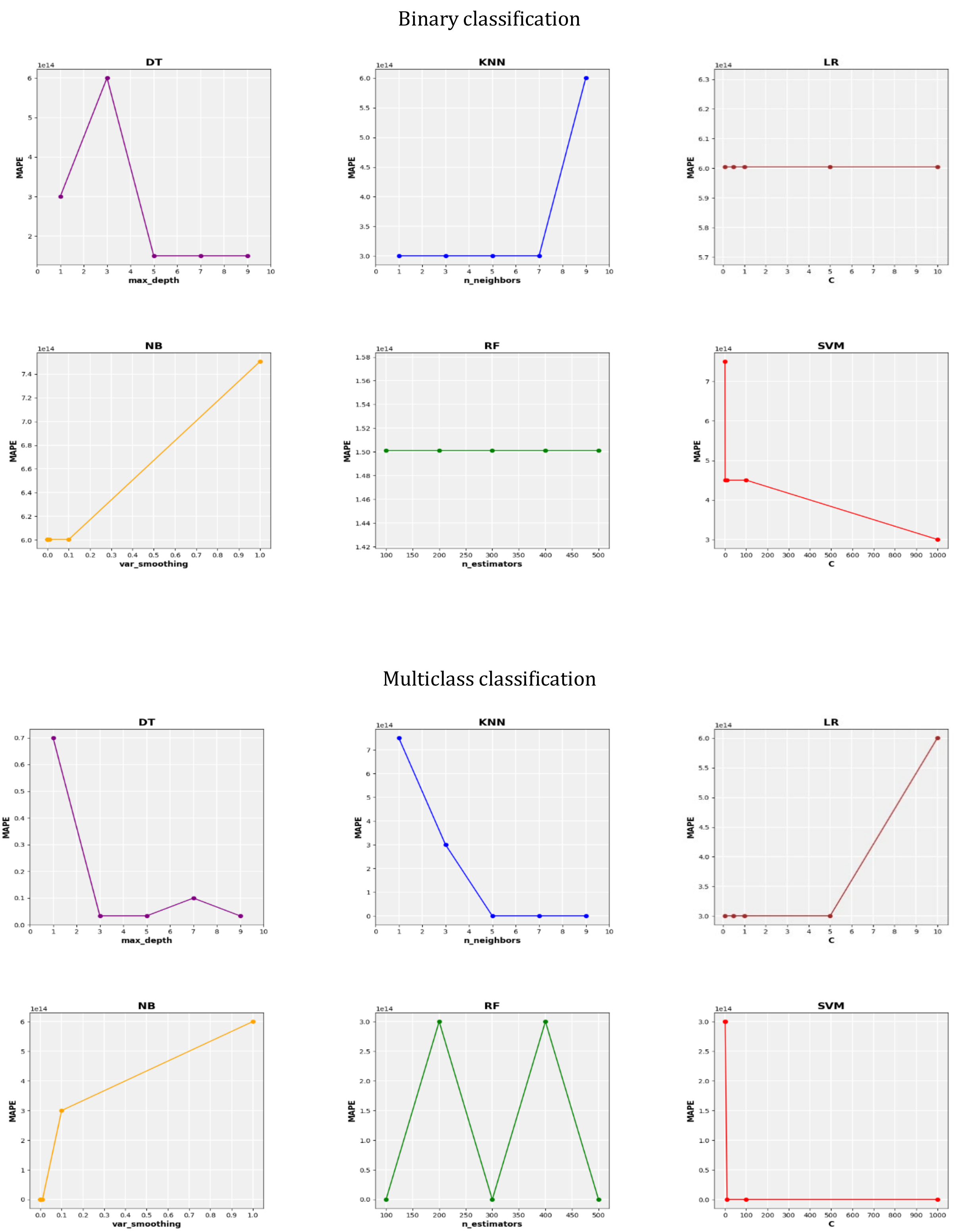

Fig. 4 *Hyperparameter tuning in binary and multiclass classification*

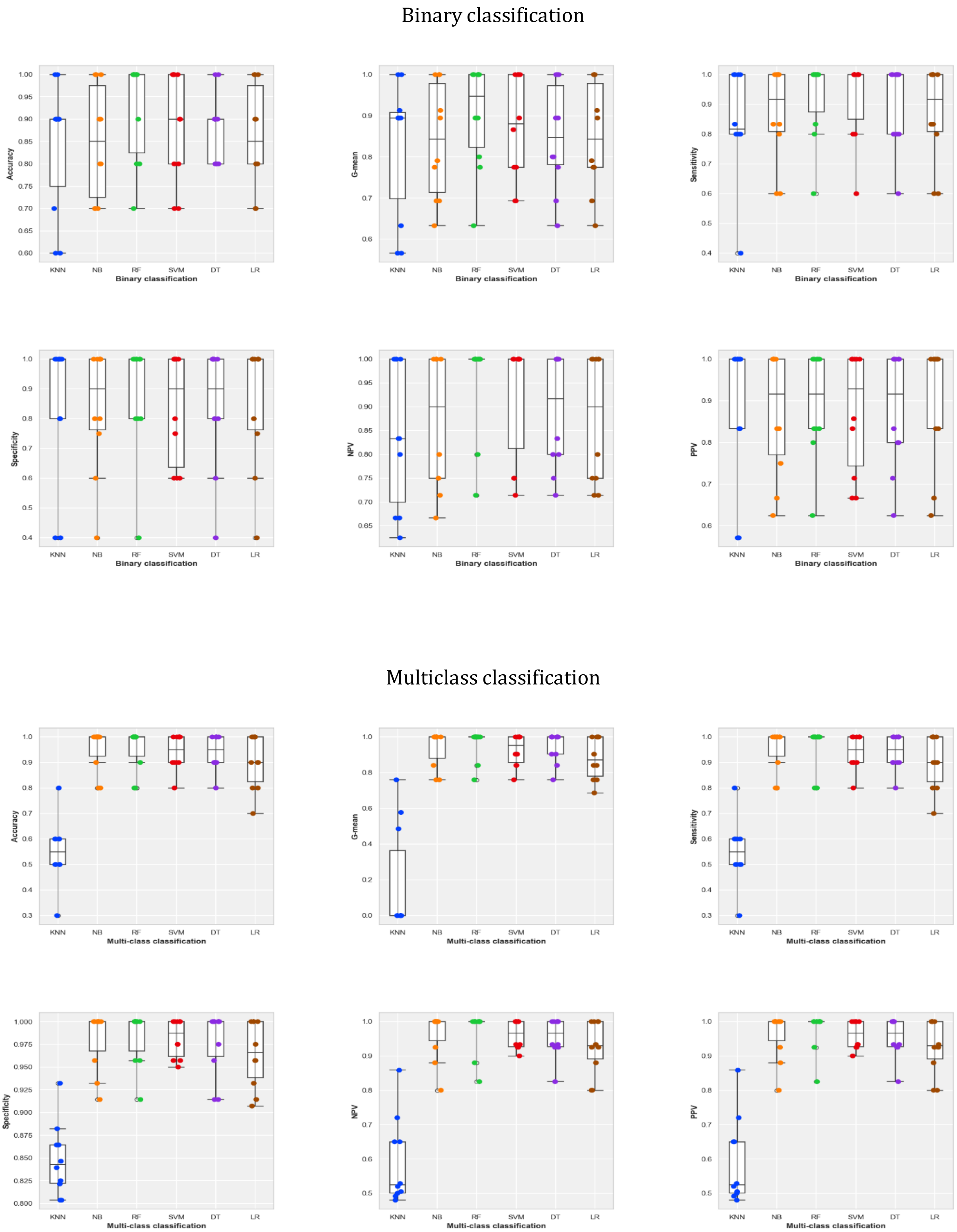

Fig. 5 *Ten-fold CV results for all classifiers in binary and multiclass classification*

5. Conclusion

The evaluation of six machine learning algorithms for classifying maternal nutritional status in Jombang District, Indonesia, demonstrates robust performance across most methods, with significant implications for public health interventions. DT, LR, NB, RF, and SVM achieved exceptional binary classification performance (AUC > 0.9), validating their utility in identifying CED risk through MUAC measurements.

These methods exhibited no statistically significant differences in accuracy ($p > 0.05$), sensitivity, or specificity during pairwise comparisons, confirming their reliability for clinical deployment in resource-constrained settings where rapid risk stratification is critical. The strong correlation between BMI and MUAC ($r = 0.92$) further supports MUAC's practicality as a simplified screening tool, particularly in rural clinics lacking advanced anthropometric measurement capabilities. However, K-nearest neighbors (KNN) underperformed in multiclass BMI categorization (AUC = 0.80), revealing inherent limitations in handling imbalanced datasets and high-dimensional feature spaces.

Several methodological constraints necessitate consideration when interpreting these findings. The study's reliance on a small, geographically limited dataset (N = 100) from a single Indonesian district compromises generalizability across diverse populations with varying socioeconomic and nutritional profiles. While manual hyperparameter tuning improved model performance—such as RF with n_estimators = 100 and SVM with C = 1000—this approach becomes unsustainable for scaling to larger datasets or real-time adaptive systems. Furthermore, the exclusion of modern ensemble methods and deep learning architectures leaves unexplored opportunities for enhanced predictive accuracy through techniques like gradient boosting or neural networks. The dataset's narrow focus on anthropometric measurements (BMI, MUAC) omits critical determinants of maternal health, including micronutrient levels, dietary patterns, and healthcare access indicators, potentially limiting the models' clinical utility.

Future research must prioritize three strategic advancements to address these limitations. First, expanding data collection through nationwide collaborations with Indonesia's District Health Offices will capture regional variability in nutritional determinants while adhering to ethical AI deployment frameworks. Second, integrating adaptive optimization techniques—such as Bayesian hyperparameter tuning and synthetic minority oversampling (SMOTE)—can automate model refinement processes for evolving datasets, ensuring sustained accuracy as healthcare systems digitize. Third, comparative studies evaluating hybrid architectures combining tree-based methods with deep learning could establish population-specific classification models while maintaining MUAC's operational advantages in low-resource clinics. These innovations must be coupled with stakeholder engagement programs to align technical capabilities with frontline healthcare workers' needs, ensuring algorithmic outputs translate into actionable clinical decisions.

Ultimately, this study underscores machine learning's transformative potential in combating maternal malnutrition within Indonesia's decentralized healthcare landscape. By bridging the gap between anthropometric screening and data-driven interventions, these methods offer a scalable pathway to achieve Sustainable Development Goal targets for maternal health. However, sustained progress demands interdisciplinary collaboration among data scientists, clinicians, and policymakers to develop context-aware systems that prioritize equity alongside technical performance metrics. The integration of ethical AI principles into maternal health programs will be paramount as Indonesia scales these technologies to address persistent disparities in healthcare access and outcomes.

Acknowledgement

This work was supported by Universitas Brawijaya under *Penelitian FILKOM Berkarya Tahun* 2024.

Conflict of Interest

The authors declare no conflict of interest.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Diva Kurnianingtyas, Nathan Daud;* ***data collection:*** *Diva Kurnianingtyas;* ***analysis and interpretation of results:*** *Diva Kurnianingtyas, Nathan Daud;* ***draft manuscript preparation:*** *Diva Kurnianingtyas, Nathan Daud. All authors (Diva Kurnianingtyas, Nathan Daud, Agus Wahyu Widodo, Tutut Herawan) reviewed the results and approved the final version of the manuscript.*

References

[1] Agustina, R., Rianda, D., Lasepa, W., Birahmatika, F. S., Stajic, V., & Mufida, R. (2023). Nutrient intakes of pregnant and lactating women in Indonesia and Malaysia: Systematic review and meta-analysis. Frontiers in Nutrition, 10, 1030343.

[2] Bachiri, Y.-A., Mouncif, H., & Bouikhalene, B. (2023). Artificial Intelligence Empowers Gamification: Optimizing Student Engagement and Learning Outcomes in E-learning and MOOCs. International Journal of Engineering Pedagogy, 13(8).

[3] Chen, E. (2022). The validation of the mid-upper arm circumference measurement in adults as a weight estimation tool in a high HIV prevalence setting. Faculty of Health Sciences.

[4] Das, A., Saimala, G., Reddy, N., Mishra, P., Giri, R., Kumar, A., Raj, A., Kumar, G., Chaturvedi, S., & Babu, S. (2020). Mid-upper arm circumference as a substitute of the body mass index for assessment of nutritional status among adult and adolescent females: learning from an impoverished Indian state. Public Health, 179, 68–75.
[5] Delnevo, G., Mancini, G., Roccetti, M., Salomoni, P., Trombini, E., & Andrei, F. (2021). The prediction of body mass index from negative affectivity through machine learning: a confirmatory study. Sensors, 21(7), 2361.
[6] El Hamdaoui, H., Boujraf, S., El Houda Chaoui, N., Alami, B., & Maaroufi, M. (2021). Improving Heart Disease Prediction Using Random Forest and AdaBoost Algorithms. International Journal of Online & Biomedical Engineering, 17(11).
[7] Endres, L. K., Straub, H., McKinney, C., Plunkett, B., Minkovitz, C. S., Schetter, C. D., Ramey, S., Wang, C., Hobel, C., & Raju, T. (2015). Postpartum weight retention risk factors and relationship to obesity at 1 year. Obstetrics & Gynecology, 125(1), 144–152.
[8] Jeffs, E., Haszard, J. J., Sharp, B., Gullam, J., & Paterson, H. (2016). Pregnant women lack accurate knowledge of their BMI and recommended gestational weight gain. The New Zealand Medical Journal (Online), 129(1439), 37.
[9] KC, A., Basel, P. L., & Singh, S. (2020). Low birth weight and its associated risk factors: Health facility-based case-control study. PloS One, 15(6), e0234907.
[10] Kpewou, D. E., Poirot, E., Berger, J., Som, S. V., Laillou, A., Belayneh, S. N., & Wieringa, F. T. (2020). Maternal mid-upper arm circumference during pregnancy and linear growth among Cambodian infants during the first months of life. Maternal & Child Nutrition, 16, e12951.
[11] Kurnianingtyas, D., & Muflikhah, L. (2024). Application of Artificial Intelligence for Maternal and Child Disorders in Indonesia: A Review. Asia Simulation Conference, 289–306.
[12] La Cruz, A., Severeyn, E., Velásquez, J., Herrera, H., & Wong, S. (2020). Assessment of Anthropometric Measurements for Obesity and Abnormal Body Fat Percentage Diagnosis Using k-means as Clustering Technique. Conference on Information and Communication Technologies of Ecuador, 177–191.
[13] Lee, B. J., Kim, K. H., Ku, B., Jang, J.-S., & Kim, J. Y. (2013). Prediction of body mass index status from voice signals based on machine learning for automated medical applications. Artificial Intelligence in Medicine, 58(1), 51–61.
[14] Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. Journal of Thoracic Oncology, 5(9), 1315–1316.
[15] Marshall, N. E., Abrams, B., Barbour, L. A., Catalano, P., Christian, P., Friedman, J. E., Hay Jr, W. W., Hernandez, T. L., Krebs, N. F., & Oken, E. (2022). The importance of nutrition in pregnancy and lactation: lifelong consequences. American Journal of Obstetrics and Gynecology, 226(5), 607–632.
[16] Munir, R. (2022). Pregnant Women's Knowledge of Chronic Energy Deficiency with Economic Status. AbdimasMu UMTAS, 1(2), 105–111.
[17] Ng, C. D., Elliott, M. R., Riosmena, F., & Cunningham, S. A. (2020a). Beyond recent BMI: BMI exposure metrics and their relationship to health. SSM-Population Health, 11, 100547.
[18] Ng, C. D., Elliott, M. R., Riosmena, F., & Cunningham, S. A. (2020b). Beyond recent BMI: BMI exposure metrics and their relationship to health. SSM-Population Health, 11, 100547.
[19] Nguyen, T. Van, Nguyen, V. D., Nguyen, T. T., Khanh, P. C. P., Nguyen, T. A., & Tran, D. T. (2020). Motorsafe: An android application for motorcyclists using decision tree algorithm. International Journal of Interactive Mobile Technologies, 14(2). https://doi.org/10.3991/ijim.v14i02.10742
[20] Novaliendry, D., Yang, C.-H., Desnelita, Y., Sanjaya, R., Lizar, Y., & Ardi, N. (2023). Hemodialysis Patient Death Prediction Using Logistic Regression. International Journal of Online & Biomedical Engineering, 19(9).
[21] Ou, J., Lin, D., & Zheng, Z. (2023). Comprehensive Evaluation of Online Experimental Teaching Quality in Colleges and Universities Based on Support Vector Machine. International Journal of Emerging Technologies in Learning (IJET), 18(12). https://doi.org/10.3991/ijet.v18i12.39697
[22] Ramadhani, C., Suksmadana, I. M. B., Ariessaputra, S., & Wijaya, I. G. P. S. (2021). MULTICLASS CLASSIFICATION OF SOUND HEALING WITH K-NEAREST NEIGHBOR ALGORITHM. DIELEKTRIKA, 8(2), 156–163.
[23] Recenti, M., Ricciardi, C., Gìslason, M., Edmunds, K., Carraro, U., & Gargiulo, P. (2020). Machine learning algorithms predict body mass index using nonlinear trimodal regression analysis from computed tomography scans. XV Mediterranean Conference on Medical and Biological Engineering and Computing–MEDICON 2019: Proceedings of MEDICON 2019, September 26-28, 2019, Coimbra, Portugal, 839–846.
[24] Salih, Y., Omar, S. M., AlHabardi, N., & Adam, I. (2023). The Mid-Upper Arm Circumference as a Substitute for Body Mass Index in the Assessment of Nutritional Status among Pregnant Women: A Cross-Sectional Study. Medicina, 59(6), 1001.
[25] Shi, Z. (2020). Improving k-nearest neighbors algorithm for imbalanced data classification. IOP Conference Series: Materials Science and Engineering, 719(1), 12072.

[26] Soliman, A., De Sanctis, V., Alaaraj, N., Ahmed, S., Alyafei, F., Hamed, N., & Soliman, N. (2021). Early and long-term consequences of nutritional stunting: from childhood to adulthood. Acta Bio Medica: Atenei Parmensis, 92(1).
[27] Suppa, R. (2023). Comparative Performance Evaluation Results of Classification Algorithm in Data Mining to Identify Types of Glass Based on Refractive Index and It's Elements. PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik, 8(1), 20–31.
[28] Trevethan, R. (2017). Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Frontiers in Public Health, 5, 307.
[29] Widodo, A. W., Kurnianingtyas, D., & Mahmudy, W. F. (2022). Optimization of Healthcare Problem using Swarm Intelligence: A Review. 2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 747–751. https://doi.org/10.1109/IEEM55944.2022.9989619
[30] Zamsuri, A., Defit, S., & Nurcahyo, G. W. (2023). Classification of Multiple Emotions in Indonesian Text Using The K-Nearest Neighbor Method. Journal of Applied Engineering and Technological Science (JAETS), 4(2), 1012–1021.