

## Performance Comparison of Logistic Regression and Neural Networks in Predicting Stunting Prevalence in Aceh Province

Alfiya Zahara<sup>1</sup>, Dahlan Abdullah<sup>2</sup>, Muhammad Daud<sup>3</sup>, Muhammad Sayuti<sup>4</sup>, Nurdin<sup>5</sup>

<sup>1</sup>Master Program in Information Technology, Faculty of Engineering, Malikussaleh University

<sup>2,3,4,5</sup>Department of Informatics, Faculty of Engineering, Malikussaleh University

### ABSTRACT

Stunting remains a significant public health problem in Indonesia, particularly in Aceh Province, where differences in prevalence across districts/cities require data-driven approaches to support targeted intervention planning. This study aims to compare the predictive performance of Logistic Regression and Neural Networks in classifying stunting prevalence across 23 districts/cities in Aceh Province. A quantitative approach was employed using secondary data from the Aceh Provincial Health Profile for 2022–2025, comprising 92 observations. The predictor variables included exclusive breastfeeding coverage, improved sanitation coverage, and vitamin A supplementation coverage, while stunting prevalence was transformed into a binary classification, with 0 representing prevalence of  $\leq 20\%$  and 1 representing prevalence of  $> 20\%$ . Logistic Regression and Multilayer Perceptron Neural Networks were implemented and evaluated using Accuracy, Precision, Recall, F1-score, and Area Under the Curve (AUC). The results showed that both models were capable of classifying stunting prevalence, while Neural Networks demonstrated better overall predictive performance. Logistic Regression achieved an Accuracy of 0.78 and an AUC of 0.82, whereas Neural Networks achieved an Accuracy of 0.89, Precision of 0.87, Recall of 0.90, and F1-score of 0.88. The findings indicate that Neural Networks provide stronger predictive performance in identifying high-stunting categories, while Logistic Regression offers greater interpretability for understanding relationships between health indicators and stunting prevalence. Therefore, Neural Networks have potential as a predictive approach for supporting the identification of high-risk areas and data-driven stunting intervention planning in Aceh Province.

**Keyword :** Stunting; Logistic Regression; Neural Networks; Machine Learning; Health Prediction; Aceh Province.

 This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

### Corresponding Author:

Muhammad Daud,  
Department of Information Technology  
Universitas Malikussaleh  
Jl. Batam, Kampus Bukit Indah, Muara Satu, Lhoseumawe, Aceh, Indonesia  
Email : mdaud@unimal.ac.id

### Article history:

Received Jun 14, 2026  
Revised Aug 17, 2026  
Accepted Sept 21, 2026

## 1. INTRODUCTION

Stunting remains a major public health problem in Indonesia and continues to require effective and data-driven intervention strategies. Stunting refers to impaired linear growth resulting from chronic nutritional deficiencies and repeated health problems during early childhood. Its consequences extend beyond physical growth and may affect cognitive development, learning capacity, immunity, productivity, and long-term health. In Aceh Province, stunting remains an important public health concern, with variations in prevalence across districts and cities. These regional differences indicate the need for analytical approaches that can identify areas with a higher probability of stunting and support more targeted intervention planning (Husain et al., 2024).

Stunting is influenced by various nutritional, health, environmental, and socioeconomic factors. In this study, three health-related indicators are considered, namely exclusive breastfeeding, improved sanitation, and vitamin A supplementation. Exclusive breastfeeding provides essential nutrients and protective components during early childhood, while adequate sanitation can reduce exposure to infections that may interfere with nutrient absorption. Vitamin A supplementation also contributes to immune function and child growth. Therefore, these indicators provide relevant information for analyzing differences in stunting prevalence across regions (Puspasari et al., 2024).

The increasing availability of health-related data has created opportunities to apply statistical and machine learning approaches to stunting analysis. Recent research published in the Journal of

Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE) has demonstrated the application of computational methods to stunting related problems. Robust Geographically Weighted Regression (RGWR) to model stunting cases among toddlers in North Sumatra. Their study developed different models for 33 districts/cities and reported a MAPE of 15%, indicating that spatial regression can be useful for estimating regional stunting cases (Siahaan & Husein, 2024). Another JCoSITTE study by Sari et al. applied the K-Means algorithm to cluster stunting cases using 395 datasets, demonstrating the use of data mining techniques to identify patterns and groups in stunting related data (Sari et al., 2023). These studies indicate that computational approaches can provide valuable support for analyzing stunting from a regional and data-driven perspective.

Research conducted by researchers from Universitas Malikussaleh has also demonstrated the application of machine learning to stunting-related problems. (Pasaribu et al., 2025) developed a Naive Bayes Classifier with Backward Elimination feature selection to classify stunting status among toddlers using data collected from the Lueng Daneun Public Health Center in Bireuen. The study highlights the potential of machine learning and feature-selection techniques for improving the classification of stunting status. (Pasaribu et al., 2025) Similarly, Junaidi, Nurdin, and Meiyanti (2025) applied the K-Nearest Neighbor (KNN) algorithm to classify child stunting status using 558 records containing variables such as age, birth weight, birth height, current weight, current height, and upper-arm circumference. The study further demonstrates the applicability of machine learning-based classification for early detection of stunting (Junaidi et al., 2025).

The application of computational methods by Universitas Malikussaleh researchers also extends to other health related prediction problems. (Putri et al., 2026) applied the Local Mean K-Nearest Neighbor algorithm to classify heart disease using medical-record data consisting of 403 observations and 10 variables. The study evaluated classification performance using accuracy, precision, recall, and F1-score, demonstrating the applicability of machine learning to complex health datasets. In addition, (Khairani Sofyan, 2018) has contributed to forecasting research through a study on export demand forecasting using the decomposition method, as documented in the official publication list of the Industrial Engineering/Logistics program at Universitas Malikussaleh. Although the research domain differs from stunting, its focus on forecasting provides a broader methodological connection to prediction based decision making.

Among the available statistical and machine learning approaches, Logistic Regression and Neural Networks provide complementary advantages. Logistic Regression is a conventional statistical method that can estimate the probability of a binary outcome and provide interpretable coefficients for examining the relationship between predictor variables and stunting status. However, its ability to capture complex nonlinear relationships may be limited. Neural Networks, in contrast, can learn nonlinear patterns and interactions among multiple variables, making them potentially suitable for prediction problems involving complex relationships. Nevertheless, Neural Networks generally provide lower interpretability than Logistic Regression. Therefore, a direct comparison between these approaches is necessary to determine whether the predictive capability of Neural Networks provides meaningful advantages over an interpretable statistical model.

Previous studies have also demonstrated differences in the performance of statistical and machine learning methods for stunting-related prediction. (Setyawan & Wakhidah, 2025) applied Logistic Regression to identify factors associated with stunting and reported an accuracy of 83%. (Wutsqa & Rezeki, 2012) developed a Multilayer Perceptron-based Neural Network for stunting prediction and obtained an accuracy of 91%. (Yuniarsyih R.A et al., 2025) compared Logistic Regression and Artificial Neural Network for predicting children's nutritional status and reported higher accuracy for Neural Network (94%) than Logistic Regression (86%), although Logistic Regression provided better interpretability. (Muthohharoh & Fatah, n.d.) further demonstrated the potential of Deep Neural Networks for stunting prediction using national-level data.

Despite these developments, several research gaps remain. Existing JCoSITTE studies have focused on spatial regression and clustering of stunting data, while studies by researchers from Universitas Malikussaleh have applied Naive Bayes and KNN for individual-level stunting classification. (Siahaan & Husein, 2024) These studies generally employ a single computational method and use individual-level or specific health-center datasets. In contrast, there remains limited research directly comparing Neural Networks and Logistic Regression using consistent nutritional and environmental indicators at the district/city level across all 23 districts/cities in Aceh Province. This regional

perspective is important because variations in health indicators across districts and cities may influence the predictive performance of different modeling approaches (Aramico & Husna, 2017).

Therefore, this study aims to compare the predictive performance of Logistic Regression and Neural Networks in classifying stunting prevalence across 23 districts/cities in Aceh Province during 2022–2025. The study uses secondary data consisting of exclusive breastfeeding coverage, improved sanitation coverage, and vitamin A supplementation coverage. Stunting prevalence is transformed into a binary classification, where 0 represents a prevalence of  $\leq 20\%$  and 1 represents a prevalence of  $> 20\%$ . Both models are evaluated using Accuracy, Precision, Recall, F1-score, and Area Under the Curve (AUC), while prediction probabilities are also generated for each district/city.

The novelty of this study lies in the direct comparative evaluation of a conventional statistical model and a machine learning model using the same nutritional and environmental indicators across all districts/cities in Aceh Province. Unlike previous studies that focus on individual stunting classification or a single prediction algorithm, this study evaluates Logistic Regression and Neural Networks under the same regional dataset and classification framework. Logistic Regression provides an interpretable baseline for understanding the relationship between health indicators and stunting prevalence, while Neural Networks provide the ability to capture potentially nonlinear patterns. The findings are expected to contribute to data-driven identification of high-risk areas and support more targeted stunting intervention and health-policy planning in Aceh Province.

## 2. RESEARCH METHOD

This study employed a quantitative research approach with a comparative predictive modeling design. The study aimed to compare the performance of Logistic Regression and Neural Networks in predicting the stunting prevalence category in Aceh Province. The research used secondary data from the Aceh Provincial Health Profile for the 2022–2025 period. The research process consisted of data acquisition, data preprocessing, descriptive analysis, model development, model testing, performance evaluation, and comparison of the two models.

### 2.1 Research Design and Data Acquisition

The data used in this study were secondary quantitative data obtained from the Aceh Provincial Health Profile for 2022–2025, published by the Aceh Provincial Health Office. The data consist of annual health indicators for all districts/cities in Aceh Province. The unit of analysis was the district/city, consisting of 23 districts/cities, observed over four years, resulting in 92 observations ( $23 \times 4$  years). (Lukmana et al., 2024).

The study used one dependent variable and three independent variables. The dependent variable was Stunting Prevalence (TB/U), while the independent variables consisted of Exclusive Breastfeeding Coverage, Improved Sanitation Coverage, and Vitamin A Supplementation Coverage. The operational definitions of the variables are presented in Table 1.

Table 1. Research Variables

| Variable                   | Type        | Operational Definition                                                                                | Unit |
|----------------------------|-------------|-------------------------------------------------------------------------------------------------------|------|
| Stunting Prevalence (TB/U) | Dependent   | Percentage of children under five with height-for-age below $-2$ SD based on the WHO growth standard  | %    |
| Exclusive Breastfeeding    | Independent | Percentage of infants aged 0–6 months receiving only breast milk without additional food or beverages | %    |
| Improved Sanitation        | Independent | Percentage of households with access to sanitation facilities that meet health requirements           | %    |
| Vitamin A Supplementation  | Independent | Percentage of children receiving vitamin A capsules according to the recommended schedule             | %    |
| Stunting Category          | Categorical | Binary classification of stunting prevalence: 0 = $\leq 20\%$ and 1 = $> 20\%$                        | %    |

The operational definitions and data sources are based on the variables used in this research. The stunting prevalence was transformed into a binary target variable for classification. Observations with a prevalence of  $\leq 20\%$  were coded as 0 (not stunting), while observations with a prevalence of

>20% were coded as 1 (stunting). This classification was applied consistently to both prediction models.

## 2.2 Data Preprocessing

Data preprocessing was conducted before model development to prepare the dataset for statistical and computational analysis. The data were collected from official sources in PDF and Excel formats, verified for consistency across years and sources, and converted into structured Excel/CSV data for processing in Python. The preprocessing stages consisted of data cleaning, data transformation, normalization, and binary target-variable formation.

Data cleaning involved checking duplicate records, correcting data-entry inconsistencies, and addressing incomplete data. Data transformation was performed by standardizing percentage values and coding categorical variables. Data normalization was applied specifically for the Neural Networks model to place the input variables on a comparable scale and support the model learning process. Finally, the continuous stunting prevalence variable was converted into the binary classification described previously.

Before prediction modeling, descriptive analysis was conducted to describe the characteristics of the dataset. The analysis included basic statistical measures such as minimum, maximum, mean, median, and standard deviation, as well as graphical presentation and initial correlation analysis. This stage was used to identify the general distribution and relationships among the health indicators before applying the prediction models.

## 2.3 Prediction Models

Two approaches were used in this study, namely Binary Logistic Regression and Neural Networks. Both models used the same three predictor variables and the same binary stunting classification to provide a consistent basis for performance comparison.

### a. Logistic Regression

Binary Logistic Regression was selected because the dependent variable consists of two categories, namely 0 ( $\leq 20\%$ ) and 1 ( $> 20\%$ ). The model estimates the probability of an observation belonging to the high-stunting category based on the three predictor variables. (Nasien et al., 2024) The logistic regression model can be expressed as:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)}} \quad (1)$$

where  $p$  represents the probability of high stunting prevalence,  $X_1$  represents Exclusive Breastfeeding,  $X_2$  represents Improved Sanitation, and  $X_3$  represents Vitamin A Supplementation. The coefficients  $\beta_0, \beta_1, \beta_2$ , and  $\beta_3$  represent the intercept and regression coefficients.

The analysis involved identifying the dependent and independent variables, assessing model suitability, estimating regression parameters using Maximum Likelihood Estimation, evaluating model performance, and interpreting the coefficients using Odds Ratios.

For classification, the predicted probability was compared with a 0.5 decision threshold. An observation with a probability greater than 0.5 was classified as stunting (1), while an observation with a probability of 0.5 or lower was classified as not stunting (0).

### b. Neural Networks

Neural Networks were implemented as a comparative machine learning model because of their ability to capture potentially complex and nonlinear relationships between the predictor variables and stunting prevalence. The model used three input variables: Exclusive Breastfeeding, Improved Sanitation, and Vitamin A Supplementation. (Ramayani Nur Hadiati et al., 2025)

The Neural Networks architecture consisted of an input layer with three input variables, one hidden layer, and an output layer with one neuron representing the stunting category. The hidden layer used the ReLU activation function, while the output layer used the Sigmoid activation function. The model was trained using the Adam optimizer and Binary Cross-Entropy loss function. The number of training epochs was determined based on model convergence.

The Neural Networks model was implemented using Python through the Anaconda environment. During the training process, the network iteratively updated its weights and biases over a number

of epochs. The output layer generated a probability value between 0 and 1 representing the predicted probability of the stunting category.

## 2.4 Training, Testing, and Model Evaluation

The dataset was divided into training and testing sets before Neural Networks modeling. The training set consisted of 74 observations (80%), while the testing set consisted of 18 observations (20%), giving a total of 92 observations. The training data were used to learn the patterns in the dataset, while the testing data were used to evaluate the model's prediction performance on previously unseen observations.

The performance of both prediction models was evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC. These metrics were selected to provide a comprehensive evaluation of classification performance. (Pratama & Baita, 2025) Accuracy measures the proportion of correctly classified observations:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

ROC-AUC was also used to evaluate the ability of the models to distinguish between the two stunting categories at different classification thresholds. The evaluation procedure compares the predicted classifications with the actual stunting categories in the testing data.

The final stage involved comparing the performance of Logistic Regression and Neural Networks using the resulting evaluation metrics. Prediction probabilities generated by the models were also used to identify districts/cities with a relatively higher probability of belonging to the high-stunting category.

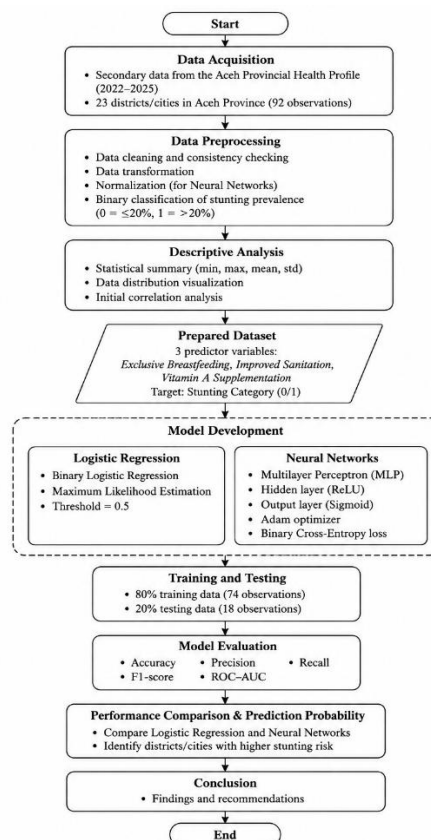


Fig 1. Research Procedure

The research procedure used in this study is presented in Figure 1. The procedure consists of several sequential stages, beginning with data acquisition and followed by data preprocessing, descriptive analysis, dataset preparation, model development, training and testing, model evaluation, performance comparison, and conclusion. The data acquisition stage involves collecting secondary data from the Aceh Provincial Health Profile for 2022–2025, covering 23 districts/cities and resulting in 92 observations. The data are then preprocessed through data cleaning, transformation, normalization for the Neural Networks model, and binary classification of stunting prevalence. Descriptive analysis is subsequently conducted to examine the characteristics and distribution of the variables. The prepared dataset is then used to develop two prediction models, namely Logistic Regression and Neural Networks. The models are trained and tested using the prepared data and evaluated based on Accuracy, Precision, Recall, F1-score, and ROC-AUC. Finally, the performance of both models is compared, and prediction probabilities are analyzed to identify districts/cities with a higher probability of stunting.

### 3. RESULTS AND DISCUSSION

This section presents the results of the analysis conducted to compare the performance of Logistic Regression and Neural Networks in predicting stunting prevalence in Aceh Province. The analysis was based on 92 observations covering 23 districts/cities during the 2022–2025 period, using exclusive breastfeeding, improved sanitation, and vitamin A supplementation as predictor variables. The results are presented through descriptive analysis, Logistic Regression modeling, Neural Networks modeling, and a comparative evaluation of both models based on their predictive performance. The findings are subsequently discussed in relation to the observed patterns and the characteristics of each modeling approach.

#### 3.1 Descriptive Analysis

The dataset consisted of 92 observations representing 23 districts/cities in Aceh Province over the 2022–2025 period. The analysis included stunting prevalence, exclusive breastfeeding coverage, improved sanitation coverage, and vitamin A supplementation coverage. The descriptive statistics of the variables are presented in Table 2.

Table 2. Descriptive Statistics of Research Variables

| Variable                  | Minimum (%) | Maximum (%) | Mean (%) |
|---------------------------|-------------|-------------|----------|
| Stunting Prevalence       | 17.4        | 36.2        | 29.5     |
| Exclusive Breastfeeding   | 51.4        | 63.0        | 57.6     |
| Improved Sanitation       | 57.0        | 70.0        | 64.2     |
| Vitamin A Supplementation | 77.6        | 86.0        | 82.1     |

As shown in Table 2, stunting prevalence varied considerably across districts/cities, ranging from 17.4% to 36.2%, with an average of 29.5%. In comparison, exclusive breastfeeding ranged from 51.4% to 63.0%, improved sanitation from 57.0% to 70.0%, and vitamin A supplementation from 77.6% to 86.0%. These results indicate substantial regional variation in stunting prevalence despite relatively stable coverage of the three health indicators. The annual trends of the four variables are presented in Table 3.

Table 3. Annual Average of Health Indicators and Stunting Prevalence

| Year | Stunting (%) | Exclusive Breastfeeding (%) | Improved Sanitation (%) | Vitamin A (%) |
|------|--------------|-----------------------------|-------------------------|---------------|
| 2022 | 32.4         | 55.4                        | 62.3                    | 81.6          |
| 2023 | 30.1         | 56.6                        | 63.7                    | 82.4          |
| 2024 | 28.7         | 57.8                        | 65.1                    | 83.2          |
| 2025 | 26.9         | 59.0                        | 67.0                    | 84.0          |

Table 3 shows a consistent decline in the average stunting prevalence from 32.4% in 2022 to 26.9% in 2025. During the same period, exclusive breastfeeding increased from 55.4% to 59.0%, improved sanitation increased from 62.3% to 67.0%, and vitamin A supplementation increased from 81.6% to 84.0%.

The decline in stunting prevalence accompanied by improvements in the three health indicators suggests a favorable temporal pattern. However, the relationship should not be interpreted solely from the annual averages because the observations vary across districts/cities. This regional variation provides the basis for further analysis using predictive models.

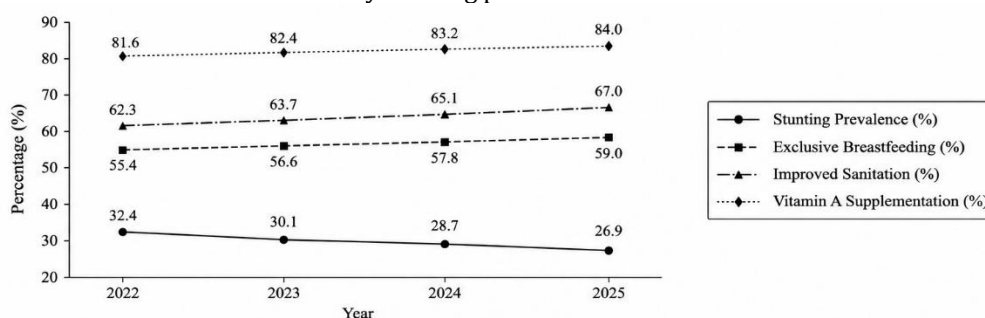


Fig 2. Trend of Stunting Prevalence in Aceh Province, 2022–2025

The trend shown in Figure 2 illustrates the gradual decline in stunting prevalence during the observation period. Nevertheless, the prevalence in 2025 remained above the 20% classification threshold used in this study, indicating that stunting remained a relevant regional health concern.

### 3.2 Logistic Regression Results

Binary Logistic Regression was applied to estimate the probability of an observation belonging to the high-stunting category. The classification threshold was set at 0.5, where a predicted probability greater than 0.5 was classified as stunting, while a probability of 0.5 or lower was classified as not stunting.

The estimated coefficients indicated negative relationships between the three health indicators and the probability of high stunting prevalence. The coefficients obtained from the model were  $-0.030$  for exclusive breastfeeding,  $-0.050$  for improved sanitation, and  $-0.040$  for vitamin A supplementation.

These results indicate that higher values of the three health indicators were associated with lower predicted probability of high stunting prevalence. Among the predictors, improved sanitation showed the largest negative coefficient, suggesting that sanitation had the strongest contribution among the three variables within the fitted Logistic Regression model. This finding is consistent with the interpretation in the thesis that environmental conditions, particularly sanitation, play an important role in stunting risk. The classification performance of Logistic Regression is presented in Table 4.

Table 4. Logistic Regression Performance

| Metric    | Value |
|-----------|-------|
| Accuracy  | 0.78  |
| Precision | 0.80  |
| Recall    | 0.75  |
| F1-score  | 0.77  |
| ROC-AUC   | 0.82  |

The model achieved an accuracy of 0.78, indicating that 78% of the observations were correctly classified. The precision of 0.80 indicates relatively good correctness in identifying the positive class, while the recall of 0.75 indicates that the model correctly identified 75% of the actual positive observations. The F1-score of 0.77 indicates a relatively balanced performance between precision and recall. The ROC-AUC of 0.82 indicates good discrimination between the two stunting categories.

The results demonstrate that Logistic Regression provides a reasonably strong baseline for stunting classification. However, its lower predictive performance compared with Neural Networks suggests that the relationship between the health indicators and stunting classification may contain nonlinear patterns that are not fully captured by the conventional logistic model.

### 3.3 Neural Networks Results

The Neural Networks model was developed using the three health indicators as input variables. The model consisted of an input layer, one hidden layer, and an output layer. The hidden layer used the ReLU activation function, while the output layer used Sigmoid activation. The model was trained using the Adam optimizer and Binary Cross-Entropy loss function.

The dataset was divided into 74 training observations (80%) and 18 testing observations (20%). The training data were used to learn the patterns in the dataset, while the testing data were used to evaluate the model on observations not used during training. The performance of the Neural Networks model is presented in Table 5.

Table 5. Neural Networks Performance

| Metric    | Value |
|-----------|-------|
| Accuracy  | 0.89  |
| Precision | 0.87  |
| Recall    | 0.90  |
| F1-score  | 0.88  |

The Neural Networks model achieved an accuracy of 0.89, which indicates that 89% of the observations were correctly classified. Its precision of 0.87 indicates that most observations predicted as high stunting were correctly classified, while the recall of 0.90 indicates a strong ability to identify observations belonging to the high-stunting category. The F1-score of 0.88 further indicates a good balance between precision and recall.

The higher recall is particularly relevant for stunting-risk identification because failing to identify a high-risk area may result in the omission of areas requiring intervention. Therefore, the Neural Networks model demonstrates stronger predictive capability for identifying the high-stunting category in this dataset.

### 3.4 Comparison of Logistic Regression and Neural Networks

The comparative performance of the two models is presented in Table 6. According to the table, Neural Networks achieved better overall performance than Logistic Regression across all four comparison metrics. The accuracy increased from 78.0% for Logistic Regression to 88.5% for Neural Networks, representing an improvement of approximately 10 percentage points. The Neural Networks model also achieved higher precision, recall, and F1-score in the comparative evaluation.

Table 6. Performance Comparison of Prediction Models

| Model               | Accuracy | Precision | Recall | F1-score |
|---------------------|----------|-----------|--------|----------|
| Logistic Regression | 0.78     | 0.77      | 0.76   | 0.76     |
| Neural Networks     | 0.89     | 0.89      | 0.87   | 0.88     |

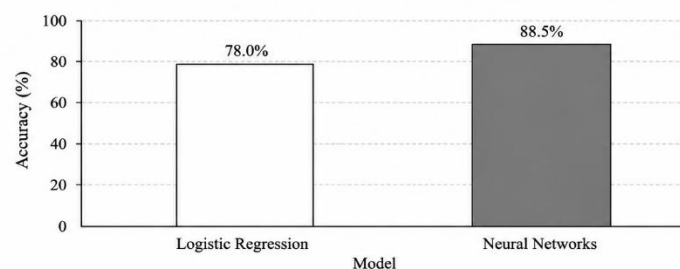


Fig 3. Comparison of Model Accuracy

Figure 3 illustrates the difference in accuracy between the two models. Neural Networks achieved an accuracy of approximately 88.5%, compared with 78.0% for Logistic Regression. The result indicates that Neural Networks were better able to capture the patterns contained in the predictor variables.

The superior performance of Neural Networks can be associated with its ability to model nonlinear relationships among the health indicators. In contrast, Logistic Regression assumes a linear relationship between the predictors and the log-odds of the outcome. The thesis identifies this

limitation as one reason why Logistic Regression may provide lower predictive performance when relationships among variables are complex. The finding is also consistent with the conceptual basis of the Neural Networks model, which is designed to identify complex and nonlinear interactions among health-related variables.

### 3.5 Discussion

The results indicate that the three health indicators show favorable temporal changes during 2022–2025, while stunting prevalence decreased over the same period. The decline from 32.4% to 26.9% suggests an improvement in the overall regional condition; however, the remaining prevalence above 20% indicates that the problem persists and that differences between districts/cities remain important.

The Logistic Regression results demonstrate that improved sanitation and vitamin A supplementation were particularly important predictors in the model, while exclusive breastfeeding also showed a protective direction. The stronger contribution of sanitation is consistent with the role of environmental health conditions in reducing repeated infections that can interfere with child nutritional status. The thesis similarly identifies sanitation and vitamin A as the factors with the most consistent influence on stunting prevalence.

The comparison between models demonstrates that the Neural Networks model outperformed Logistic Regression. This result supports the expectation that nonlinear machine learning approaches may provide advantages when relationships between predictors and outcomes are complex. Neural Networks are able to learn interactions among input variables during training, whereas Logistic Regression provides a more constrained functional relationship.

The finding is also consistent with previous research cited in the Introduction. Previous studies have applied different computational approaches to stunting, including RGWR for regional modeling and K-Means for clustering, while studies by researchers from Universitas Malikussaleh have applied Naive Bayes and KNN for stunting classification. The present study extends these approaches by directly comparing a conventional statistical classifier, Logistic Regression, with a Neural Networks model using the same regional dataset and predictor variables across all 23 districts/cities in Aceh.

However, the results should be interpreted with caution because the dataset contains only 92 observations and represents aggregated district/city-level data. The thesis explicitly notes that the relatively small and aggregated dataset may affect the stability of the Neural Networks model. Therefore, although Neural Networks achieved higher predictive performance in this study, the result should not be interpreted as evidence that Neural Networks are universally superior to Logistic Regression. Rather, the finding indicates that Neural Networks provided better predictive performance under the specific data and modeling conditions of this study.

From a practical perspective, the results support the use of predictive modeling as a supplementary tool for identifying areas with higher stunting risk. Neural Networks can be used when predictive accuracy is prioritized, while Logistic Regression remains valuable when interpretability is important for explaining relationships between health indicators and stunting prevalence. This complementary use is consistent with the thesis conclusion that Logistic Regression is more interpretable, whereas Neural Networks provide stronger predictive performance.

## 4. CONCLUSION

This study compared the performance of Logistic Regression and Neural Networks in classifying stunting prevalence across 23 districts/cities in Aceh Province using data from 2022–2025. The results indicate that both methods were able to classify stunting prevalence, but they provide different advantages. Logistic Regression offers greater interpretability in explaining the relationship between health determinants and stunting status, making it useful as a statistical baseline and for policy-oriented analysis. In contrast, Neural Networks are more adaptive in identifying complex and nonlinear relationships among the predictor variables and are therefore more suitable when predictive performance is prioritized.

Based on the evaluation results, Neural Networks demonstrated superior predictive performance compared with Logistic Regression. Logistic Regression achieved an accuracy of 0.78 and an AUC of 0.82, while Neural Networks achieved an accuracy of 0.89, precision of 0.87, recall of 0.90, and F1-score of 0.88. These results indicate that Neural Networks provided stronger classification performance in identifying districts/cities with high stunting prevalence in the study dataset. However,

the relatively small and aggregated dataset should be considered when interpreting the stability and generalizability of the Neural Networks model.

The analysis also indicates that improved sanitation, vitamin A supplementation, and exclusive breastfeeding are relevant health indicators associated with stunting prevalence. The findings highlight the importance of improving sanitation access and strengthening nutritional interventions as part of efforts to reduce stunting in Aceh Province. Overall, the findings demonstrate that Neural Networks can serve as a predictive approach for supporting data-driven identification of high-risk areas, while Logistic Regression remains valuable for providing interpretable information to support health policy and intervention planning.

## ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Universitas Malikussaleh for the academic support provided during this research. The authors also gratefully acknowledge the Aceh Provincial Health Office for providing the health profile data used in this study. The authors further appreciate all parties who contributed to the completion of this research and manuscript.

## REFERENCES

- Aramico, B., & Husna, Z. (2017). Analisis determinan stunting pada baduta di Wilayah Kerja Puskesmas tahun 2016. *Jurnal Gizi Dan Dietetik Indonesia (Indonesian Journal of Nutrition and Dietetics)*, 4(3), 154. [https://doi.org/10.21927/ijnd.2016.4\(3\).154-160](https://doi.org/10.21927/ijnd.2016.4(3).154-160)
- Husain, H., Fatma Dewi, A., & Wardani, A. E. (2024). Pemodelan Prevalensi Stunting Indonesia Menggunakan Regresi Nonparametrik Spline Truncated Modeling of Indonesian Stunting Prevalence Using Spline Truncated Non-parametric Regression. In *Journal of Analytical Research* (Vol. 3, Number 1).
- Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases. (2023). *Journal of Computer Science, Information Technology and Telecommunication Engineering*, 4(2). <https://doi.org/10.30596/jcositte.v4i2.15765>
- Junaidi, A., Nurdin, N., & Meiyanti, R. (2025). Klasifikasi Status Anak Stunting Menggunakan Metode K-Nearest Neighbor. *Rabit: Jurnal Teknologi Dan Sistem Informasi Univrab*, 10(2), 1435–1445. <https://doi.org/10.36341/rabit.v10i2.6536>
- Khairani Sofyan, D. (2018). *Forecasting Of Optimal Raw Material Supplies For Paving Block Making In UD. Karya Jaya*.
- Lukmana, H. H., Al-Husaini, M., Puspareni, L. D., & Hoeronis, I. (2024). Rancang Bangun Sistem Informasi Deteksi Dini Stunting dengan Metode Artificial Neural Network. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, 12(3), 565. <https://doi.org/10.26418/justin.v12i3.80119>
- Muthohharoh, N., & Fatah, Z. (n.d.). *Analisis Perbandingan Algoritma Decision Tree dan Deep Learning untuk Klasifikasi Balita Stunting Menggunakan Rapidminer*. 2, 2024. Retrieved <https://www.kaggle.com/datasets/rendiputra/stunting-balita-detection-121k>
- Nasien, D., Darwin, R., Cia, A., Leo Winata, A., Go, J., Charles Wijaya, R., & Charles Lo, K. (2024). *Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, Dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes* (Vol. 4).
- Pasaribu, H. M. S., Abdullah, D., & Rosnita, L. (2025). Classification of Stunting Status Using the Naive Bayes Classifier Algorithm with Backward Elimination Feature Selection. *JINAV: Journal of Information and Visualization*, 6(1), 17–25. <https://doi.org/10.35877/454ri.jinav4100>
- Pratama, R. Y., & Baita, A. (2025). Prediksi Stunting pada Anak Balita Menggunakan Algoritma Extreme Gradient Boosting dan Bayesian Optimization. *Journal of Technology and Informatics (JoTI)*, 7(2), 175–189. <https://doi.org/10.37802/joti.v7i2.1174>
- Puspasari, C., Mardhiah, A., Husniati, A. M., Aini, N. Q., & Apriliani, D. (n.d.). *The Reception of Health Information to Reduction Stunting*.
- Putri, A. Z., Daud, M., Al, H., & Aidilof, K. (2026). Klasifikasi Penyakit Jantung Berbasis Data Rekam Medis Menggunakan Algoritma Local Mean K-Nearest Neighbor. *Jurnal Nasional Teknologi dan Sistem Informasi* 12(01), 134–143. <https://doi.org/10.25077/TEKNOSI.v12i1.2026.134-143>
- Ramayani Nur Hadiati, Wanti Perinduri Sihotang, Ario Surya Trinata, & Bunga Mardhotillah. (2025). Analisis dan Upaya Menekan Faktor-Faktor Penyebab Stunting Menggunakan Neural Network Regression Guna Mencapai SDGs di Provinsi Jambi. *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan*, 3(4), 4237–4242. <https://doi.org/10.31004/jerkin.v3i4.1229>
- Robust Geographically Weighted Regression Modeling In Cases Of Stunting Toddlers In North Sumatera Utara. (2024). *Journal of Computer Science, Information Technology and Telecommunication Engineering*. <https://doi.org/10.30596/jcositte.v5i2.20934>

- Setyawan, N. H., & Wakhidah, N. (2025). Analisis Perbandingan Metode Logistic Regression, Random Forest, Gradient Boosting untuk Prediksi Diabetes. *JIPi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 10(1), 150–162. <https://doi.org/10.29100/jipi.v10i1.5743>
- Wutsqa, D. U., & Rezeki, S. (2012). Perbandingan Model Neural Network dan Regresi Logistik pada Kasus Masa Studi Mahasiswa Jurusan Pendidikan Matematika FMIPA UNY. *PYTHAGORAS Jurnal Pendidikan Matematika*, 4(1). <https://doi.org/10.21831/pg.v4i1.683>
- Yuniarsyih R.A, R. D., Muhadi, R. A., Fitrianto, A., & Silvianti, P. (2025). Analisis Regresi Logistik Biner dan Random Forest untuk Prediksi Faktor-Faktor Stunting di Pulau Jawa. *Euler: Jurnal Ilmiah Matematika, Sains Dan Teknologi*, 13(2), 147–156. <https://doi.org/10.37905/euler.v13i2.31680>

## BIOGRAPHIES OF AUTHORS

|                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | <p><b>Alfiya Zahara</b> is a graduate of the Master of Information Technology Program at Universitas Malikussaleh, Aceh, Indonesia. She received her Bachelor's degree in Informatics from Politeknik Negeri Lhokseumawe. Her research interests include information technology, data mining, machine learning, artificial intelligence, and data analysis. Her research focuses on the application of computational and statistical methods to address problems in public health and other data driven fields. She is also interested in information systems development and the implementation of technology to support decision-making. She can be contacted at email: <a href="mailto:alfiya.227110201019@mhs.unimal.ac.id">alfiya.227110201019@mhs.unimal.ac.id</a>.</p>                                                                   |
|   | <p><b>Dahlan Abdullah</b> is a Professor at the Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia. He is a senior lecturer and researcher in the fields of Artificial Intelligence, Computer Networks, Information Systems, Optimization, and Web-Based Applications. He has been actively involved in the development of Informatics and Information Systems programs at Universitas Malikussaleh, including the Master Program of Information Technology. His research interests focus on artificial intelligence, information systems, computer networks, optimization, and web-based applications. He can be contacted at email: <a href="mailto:dahlan@unimal.ac.id">dahlan@unimal.ac.id</a>.</p>                                                                                                                          |
|  | <p><b>Muhammad Daud</b> is an Associate Professor at the Faculty of Engineering at Malikussaleh University, Aceh, Indonesia. He works as a senior lecturer and researcher in the Undergraduate Program of Electrical Engineering, Master Program of Renewable Energy Engineering, and Master Program of Information Technology. His research interests include renewable energy systems, hybrid renewable power generation, power system optimization, energy management, and sustainable electrical energy technologies. Apart from teaching, he is also active as Co-Chief Editor at the Journal of Renewable Energy, Electrical, and Computer Engineering (JREECE). Currently, as Dean of the Faculty of Engineering, Malikussaleh University. He can be contacted at email: <a href="mailto:mdaud@unimal.ac.id">mdaud@unimal.ac.id</a>.</p> |
|  | <p><b>Muhammad Sayuti</b> is Professor at the Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia. He is a senior lecturer and researcher in the Undergraduate Program of Industrial Engineering and the Master Program of Renewable Energy Engineering. His research interests include materials, renewable energy, and industrial management. He is also active as a reviewer for the Journal of Renewable Energy, Electrical, and Computer Engineering (JREECE). He can be contacted at email: <a href="mailto:sayuti_m@unimal.ac.id">sayuti_m@unimal.ac.id</a>.</p>                                                                                                                                                                                                                                                           |

|                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <p><b>Nurdin</b> is a Professor at Faculty of Engineering at Malikussaleh University, Aceh, Indonesia. He works as senior lecturer and researcher at Undergraduate Program of Informatics and Master Program of Information Technology. His research interests include artificial intelligence, data mining, and information technology. In addition to teaching and research activities, he is actively involved as an editor and reviewer for several scientific journals. He can be contacted at <a href="mailto:nurdin@unimal.ac.id">nurdin@unimal.ac.id</a>.</p> |
|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|