Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.v10i1.15434
e-ISSN : 2541-2019
p-ISSN : 2541-044X

A Hybrid SOM-LSTM Model for Childhood Stunting Risk Classification: A Case Study in a Maternity Hospital in Medan

Frans Mikael Sinaga[1*)], Ferawaty[2)], Tetti Sinaga[3)]
[1,2,3)] Universitas Pelita Harapan, Indonesia
[1*)] frans.sinaga@uph.edu, [2)]ferawaty.fik@uph.edu, [3)]tetti.sinaga@uph.edu

Submitted :Oct 16, 2025 | **Accepted** : Nov 16, 2025 | **Published** : Feb 25,2026

Abstract: Childhood stunting presents a persistent global and national health challenge, significantly impacting human resource quality. Clinical settings, particularly maternity hospitals, records a high volume of stunting cases, generating large-scale health data that necessitates advanced analytical methods for accurate risk identification. This paper proposes an integrated Self-Organizing Map (SOM)-Long Short-Term Memory (LSTM) model for robust stunting risk classification. SOM effectively clusters diverse patient data based on characteristic similarities, while LSTM leverages historical health records to predict future stunting probability. This approach aims to enhance predictive accuracy and facilitate proactive medical intervention. Simulation and evaluation results demonstrate that the proposed SOM-LSTM model significantly improves the reliability of early stunting risk detection compared to conventional methods, contributing to data-driven decision-making and advancing AI-based hospital transformation.

Keywords: Artificial Intelligence, Big Data, Long Short-Term Memory, Self-Organizing Map, Stunting

INTRODUCTION

Nutrition serves as the fundamental cornerstone for developing high-quality human resources and sustaining national development. Conversely, poor nutritional status directly impacts the quality of health, competitiveness, and productivity of future generations. Recognizing this critical link, Sustainable Development Goal (SDG) Target 2.2 explicitly aims to eradicate all forms of malnutrition, which includes fulfilling the nutritional needs for vulnerable populations such as adolescent girls, pregnant and lactating mothers, and the elderly. Among the various forms of malnutrition, stunting is a global concern. It is defined as a condition of growth failure in toddlers resulting from chronic malnutrition that commences during gestation and continues until two years of age (World Health Organization, 2025). The consequences of stunting extend beyond physical growth, leading to long-term detrimental effects on cognitive development, intelligence, and productivity in adulthood.

In Indonesia, stunting continues to pose a significant public health challenge. The results of the 2023 Indonesian Health Survey (SKI) indicated only a marginal decrease in national stunting prevalence, moving from 21.6% in 2022 to 21.5% in 2023 (Kementerian Kesehatan Republik Indonesia., 2023). This current achievement remains considerably distant from the ambitious target stipulated in the National Medium-Term Development Plan (RPJMN) 2020–2024, which seeks to reduce stunting prevalence to 14% by 2024 (Bappenas, 2019).

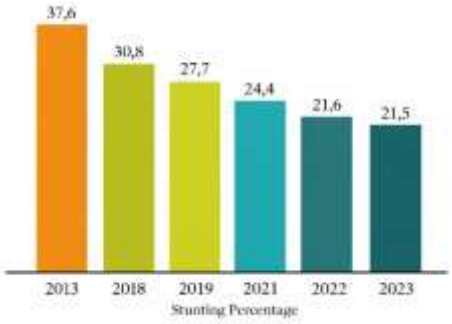

Figure 1. Flowchart of Stunting Identification System

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.v10i1.15434
e-ISSN : 2541-2019
p-ISSN : 2541-044X

To address the urgency of this pressing issue, the Indonesian government has formulated the National Strategy (Stranas) for the Acceleration of Stunting Prevention (Stunting) as an integrated guideline for all stakeholders, encompassing central, regional, and cross-sectoral levels. This strategy emphasizes five core pillars for accelerating stunting management (Sekretariat Wakil Presiden Republik Indonesia, 2021):

1. Enhancing commitment and leadership vision.
2. Promoting behavior change, communication, and community empowerment.
3. Ensuring the convergence of specific and sensitive interventions.
4. Strengthening food and nutrition security.
5. Fortifying data, information, research, and innovation systems.

The implementation of data-driven and artificial intelligence (AI)-based technology is emerging as a strategic approach to reinforce the fifth pillar of this National Strategy. The continuously growing volume of child health data, sourced from hospitals and other healthcare facilities, can be effectively leveraged as big data to enable more accurate and personalized detection and prediction of stunting risk. The central problem addressed in this research revolves around the development of an accurate and efficient classification model designed to identify and predict stunting risk among children, leveraging the extensive health big data available at a clinical maternity setting. Given the multifaceted nature of stunting's causative factors, which include detailed data on height, weight, dietary patterns, nutritional status, and socioeconomic conditions, an analytical approach rooted in artificial intelligence is deemed essential. Furthermore, the development of this model is anticipated to reinforce Pillar 5 of the National Strategy for the Acceleration of Stunting Prevention, specifically the advancement and strengthening of systems, data, information, research, and innovation, through the strategic utilization of technology to support data-driven decision-making (Ndagijimana et al., 2023; Sinaga et al., 2023).

RELATED WORK ON STUNTING RISK CLASSIFICATION

Effective stunting prevention and intervention strategies critically rely on accurate risk classification and prediction. This section reviews existing literature pertinent to stunting risk classification, encompassing traditional and machine learning approaches, along with detailed discussions on the individual and combined applications of Self-Organizing Map (SOM) and Long Short-Term Memory (LSTM) models (Ndagijimana et al., 2023; Novalina et al., 2025). This review aims to establish the current state of the art and highlight the novelty of the proposed hybrid approach within the context of stunting risk assessment (Guo et al., 2024; Ma et al., 2022; Sinaga et al., 2023).

Stunting Prediction and Classification

Historically, stunting risk assessment has often relied on traditional statistical methods and manual anthropometric measurements, as emphasized by global health organizations (World Health Organization, 2021). While these methods provide foundational insights, their ability to handle large, complex, and dynamic datasets, especially those influenced by a multitude of interconnected factors, is limited. More recent efforts have focused on leveraging advanced analytical techniques to enhance the accuracy and efficiency of stunting detection (Ndagijimana et al., 2023; Novalina et al., 2025).

Machine learning (ML) algorithms have emerged as powerful tools for predicting and classifying health conditions, including malnutrition and stunting. Studies have explored various ML classifiers such as logistic regression, k-nearest neighbor (KNN), Support Vector Machines (SVM), decision trees, Random Forest (RF), and Extreme Gradient Boosting (XGBoost) for stunting risk prediction (Sugihartono Tri, Wijaya Benny, Marini, Alkayes Ahmad Faqih, 2025). These works demonstrate the potential of data-driven models to identify at-risk populations based on diverse factors like anthropometric data, socioeconomic status, and maternal health. For instance, research in Indonesia has benchmarked several ML algorithms for stunting risk prediction using national survey data, achieving accuracies above 75% (Novalina et al., 2025). Comparative studies, including one focusing on optimizing stunting detection with SMOTE and various ML classifiers (XGBoost, RF, SVM, k-NN), have shown significant improvements in performance, especially in recall and ROC-AUC metrics, which are critical for minimizing false negatives in healthcare settings (Sugihartono Tri, Wijaya Benny, Marini, Alkayes Ahmad Faqih, 2025). Another study utilized ML techniques to construct an accurate model for early stunting detection in Rwandan children, identifying key predictive attributes such as mother's height, child's age, and parental education.

The integration of AI, including ML, is recognized as a potential solution to improve early stunting detection processes, capable of developing predictive models based on anthropometric data and other risk factors with higher accuracy. Efforts are also exploring sensor-based anthropometric equipment combined with machine learning for streamlined stunting detection, achieving high accuracy rates in classification. Despite these advancements, a comprehensive and robust approach that fully leverages complex, time-dependent health big data for nuanced stunting risk classification and prediction remains an area of active research (Guo et al., 2024; Sinaga et al., 2023).

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Applications of Self-Organizing Map (SOM) in Healthcare/Data Clustering

The Self-Organizing Map (SOM), introduced by Teuvo Kohonen, is a powerful unsupervised learning-based artificial neural network designed to project high-dimensional data into a lower-dimensional (typically two-dimensional) space while preserving the topological relationships of the input data (Kohonen, 2013). This characteristic makes SOM particularly effective in detecting hidden patterns and structures within complex and large datasets, without the need for pre-labeled data (Han et al., 2019).

In healthcare and data clustering, SOM has been successfully applied to uncover intricate patterns in vast datasets. Its ability to cluster data based on inherent similarities makes it well-suited for scenarios involving numerous variables, such as those encountered in child health data including height, weight, nutritional status, and various socioeconomic factors. By visualizing these clusters, SOM facilitates easier interpretation of complex patient profiles by medical professionals, aiding in targeted decision-making. This includes applications in patient stratification, disease pattern identification, and general data exploration where explicit labels are scarce or relationships are not immediately apparent (Sinaga et al., 2023)The two-dimensional map output of SOM provides an intuitive visualization that can highlight distinct groups with shared characteristics, which is invaluable for identifying specific risk patterns in a multidimensional health context (Guo et al., 2024).

Applications of Long Short-Term Memory (LSTM) in Prediction/Time-Series Analysis

Long Short-Term Memory (LSTM) networks are a specialized type of Recurrent Neural Network (RNN) particularly adept at handling and learning from sequential or time-series data. Unlike traditional neural networks, LSTMs are equipped with internal memory cells and sophisticated gating mechanisms (input, forget, and output gates) that enable them to capture long-term dependencies and remember information over extended periods, effectively mitigating the vanishing gradient problem common in standard RNNs.

In the realm of health prediction and time-series analysis, LSTM models have demonstrated superior performance in diverse applications. Their capacity to analyze chronological data makes them highly relevant for predicting future health outcomes based on a patient's historical records. For instance, LSTMs have been employed for stock price prediction involving time-series data, showcasing their ability to model complex temporal patterns (Greff et al., 2017; Wei et al., 2023). In health, this translates to analyzing a child's periodic medical history—including sequential measurements of weight, height, food consumption, and immunization records—to predict the potential risk of stunting based on developmental trends over time (Sinaga et al., 2023). Research highlights LSTM's effectiveness in classifying and predicting children's nutritional status by leveraging temporal features, consistently outperforming traditional baseline models in terms of accuracy and precision.

Hybrid SOM-LSTM Models

The integration of different artificial intelligence models, particularly those combining the strengths of clustering and sequential prediction, has gained traction in addressing complex real-world problems. Hybrid models, such as those combining SOM and LSTM, leverage the complementary capabilities of each component to achieve enhanced performance that neither could achieve alone. SOM's strength in identifying latent patterns and structuring complex, high-dimensional data, when combined with LSTM's proficiency in learning from and predicting based on time-series data, offers a powerful synergistic approach (Guo et al., 2024; Sinaga et al., 2023a).

Prior studies have explored hybrid SOM-LSTM models in various domains, demonstrating their superior performance in classification and prediction tasks. Examples include groundwater level prediction in karst critical zone aquifers and load data recovery methods, where the hybrid approach improved accuracy by considering both spatial patterns and temporal dynamics. Another relevant application involves sentiment analysis, where SOM and LSTM algorithms are combined to process and classify textual data, highlighting the versatility of this hybrid architecture across different data types. Researchers have also explored hybrid machine learning frameworks for stunting classification that integrate clustering methods with boosting algorithms, yielding high accuracy (Ndagijimana et al., 2023; Novalina et al., 2025).

While the individual strengths of SOM for data clustering and LSTM for time-series prediction are well-established, and hybrid models have been explored, the specific application and optimization of an integrated SOM-LSTM framework for granular stunting risk classification utilizing comprehensive big health data, particularly within the context of a maternity hospital, represents a novel contribution. This research aims to develop a tailored hybrid model that not only accurately identifies stunting risk patterns through SOM-driven clustering but also predicts future occurrences by leveraging the temporal insights from LSTM, providing a robust, data-driven decision support system to accelerate stunting prevention efforts. The novelty lies in the specific fusion of these powerful techniques for this particular public health challenge, with a focus on real-world hospital data and actionable clinical insights towards AI-based hospital transformation.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.v10i1.15434
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

This research leverages an artificial intelligence-based method, specifically the integration of Self-Organizing Map (SOM) and Long Short-Term Memory (LSTM), for the classification of stunting risk among children at a clinical maternity setting (Guo et al., 2024; Sinaga et al., 2023). This methodology is designed to effectively manage complex and multidimensional big health data, which includes a variety of variables such as height, weight, dietary patterns, nutritional status, and socioeconomic conditions. The sequential stages of this research are outlined in Fig. 2.

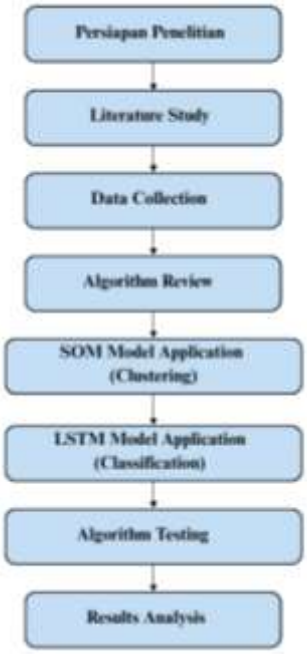

Fig. 2 Flowchart Diagram

Data Collection and Pre-processing

Data for this research were obtained as secondary data from a maternity hospital in Medan, Indonesia, which routinely records anthropometric and clinical information of newborns and children aged 0–59 months. The dataset used in this study consists of more than 300 child health records collected between 2021 and 2025, encompassing a complete set of medical, nutritional, and sociodemographic attributes relevant to stunting assessment. These attributes include body weight (kg), body height or length (cm), age (months), sex, weight-for-age z-score, height-for-age z-score, disease history, parental education level, maternal BMI, household income, immunization completeness, and feeding pattern scores. This contextual dataset reflects real operational conditions of pediatric monitoring and supports the development of clinically relevant predictive models.

Following data acquisition, a comprehensive pre-processing procedure was conducted to ensure readiness for modeling. The steps included data cleaning to address duplicates, inconsistent formats, and missing values; normalization of numerical variables using Min–Max scaling; and encoding of categorical attributes using one-hot encoding. The processed dataset was then partitioned into three subsets to support unbiased model evaluation: 70% for training, 15% for validation, and 15% for testing. The resulting dataset contains 511 complete records across 12 key attributes, forming a clean and structured input suitable for subsequent SOM clustering and LSTM prediction stages.

Self-Organizing Map (SOM) for Data Clustering

The Self-Organizing Map (SOM) is employed to cluster the collected child data into distinct stunting risk categories based on shared characteristics. As an unsupervised learning method, SOM is uniquely capable of identifying hidden patterns within the data without requiring prior labeling (Kohonen, 2013; Surucu et al., 2023). The resulting clusters will be visually represented to facilitate intuitive interpretation by the medical team. The primary output of this stage will be a mapping of clusters that delineate varying levels of stunting risk based on individual child profiles. Successful execution is measured by the clarity and medical-operational interpretability of the identified clusters (Griffin et al., 2022; Sugihartono Tri, Wijaya Benny, Marini, Alkayes Ahmad Faqih, 2025).

The initial step in the SOM process involves the initialization of weights for each neuron. These weights are randomly assigned values within the range of the input data:Following data collection, the pre-processing stage is crucial for data quality assurance prior to its utilization in modeling (Guo et al., 2024). This involves several key

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.v10i1.15434
e-ISSN : 2541-2019
p-ISSN : 2541-044X

steps: data cleaning to address anomalies, duplications, and missing values; normalization to standardize data scales; and encoding of categorical variables into a numerical format. The thoroughly processed dataset is then partitioned into distinct training and testing sets. The output of this stage is a clean, structured, and ready-to-use dataset for subsequent modeling processes.

Eq. (1) – Inisialisasi Bobot SOM

$$w_i = \text{random}(x) \quad (1)$$

After the input data is provided, the next step is to find the neuron that is closest to the input vector, called the Best Matching Unit (BMU). The BMU is determined by calculating the Euclidean distance between the input vector and each neuron's weight:

Eq. (2) — Best Matching Unit (BMU)

$$\text{BMU} = \arg\min_i \|x - w_i\| \quad (2)$$

Once the BMU is determined, all neuron weights, especially those around the BMU, are updated to move closer to the input vector. This process follows the equation:

Eq. (3) — Update Bobot SOM

$$w_j(t+1) = w_j(t) + \alpha(t) \cdot h_{ci}(t) \cdot (x(t) - w_j(t)) \quad (3)$$

The neighborhood function $h_{ci}(t)$ is used to determine the extent of the BMU's influence on its neighboring neurons, calculated using a Gaussian function:

Eq. (4) — Gaussian Neighborhood Function

$$h_{ci}(t) = \exp\left(-\frac{\|r_j - r_c\|^2}{2\sigma^2(t)}\right) \quad (4)$$

The final result of SOM application is a two-dimensional map that visualizes clusters based on the similarity of child data characteristics.

Long Short-Term Memory (LSTM) for Stunting Prediction

Following the formation of clusters via SOM, the Long Short-Term Memory (LSTM) model is applied as a predictive classification tool. Operating within a supervised learning framework, LSTM leverages historical data to forecast the probability of a child experiencing stunting within a defined timeframe. The principal advantage of LSTM lies in its inherent capability to retain long-term information and effectively manage the temporal dynamics embedded within sequential data (Tang et al., 2022; Waqas & Humphries, 2024).

The architecture of LSTM is composed of several core components known as gates. These gates regulate the flow of information, controlling what data is retained, discarded, or propagated from one step to the next. Each gate plays a distinct role in this process. The initial gate within the LSTM structure is the forget gate, which determines what information from the preceding memory needs to be disregarded:

Eq. (5) — Forget Gate

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$

Subsequent to the determination of information to be forgotten, the next step involves deciding what new information will be stored in the cell memory. This phase is executed through two distinct components:

a. Input Gate:

Eq. (6) — Input Gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (6)$$

b. Candidate Memory:

Eq. (7) — Candidate Memory

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (7)$$

The cell memory is then updated by integrating the retained old information (preserved by the forget gate) with the newly filtered information (processed by the input gate):

Eq. (8) — Update Cell State

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (8)$$

Finally, a portion of the cell memory is utilized to generate the output (hidden state) at time t, managed by the output gate:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.v10i1.15434
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Eq. (9) — Output Gate LSTM

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$

Eq. (10) — Hidden State LSTM

$$h_t = o_t \odot \tanh(C_t) \quad (10)$$

The LSTM model ultimately yields a predictive classification regarding the likelihood of a child experiencing stunting within a specified time horizon.

Model Integration and System Architecture

The integration of the Self-Organizing Map (SOM) and Long Short-Term Memory (LSTM) models forms a unified two-stage architecture designed to classify stunting risk based on multidimensional clinical and anthropometric child data. In the first stage, SOM performs unsupervised clustering to identify latent structures within the dataset, producing topological groupings that represent similarities in nutritional status, sociodemographic background, and health characteristics (Guo et al., 2024; Kohonen, 2013).

In the second stage, the cluster-enhanced feature vectors generated by SOM are fed into an LSTM-based classifier. While LSTM is traditionally employed for sequential modeling, its gated architecture can also function as a deep nonlinear classifier capable of learning complex relationships between clinical variables in a static input setting (Waqas & Humphries, 2024).By integrating SOM's unsupervised feature structuring with LSTM's capacity for high-level representation learning, the hybrid system improves classification robustness and reduces misclassification rates, particularly for children near stunting risk thresholds. Hyperparameter tuning was applied separately to both SOM and LSTM components to ensure optimal performance and generalization across the dataset (Sinaga et al., 2023). This integrated architecture ensures that multidimensional health data are effectively transformed, structured, and analyzed within a coherent modeling pipeline, resulting in a predictive system that aligns with the operational needs of pediatric risk assessment in clinical settings.

Model Configuration

The configuration of the Self-Organizing Map (SOM) and Long Short-Term Memory (LSTM) models is presented in this subsection to ensure full reproducibility and methodological transparency. All hyperparameters, training configurations, and computational environments are explicitly documented according to good scientific practice (Kohonen, 2013; Greff et al., 2017).

1. Self-Organizing Map (SOM) Configuration

The SOM was utilized as an unsupervised clustering technique to uncover latent structures within the 12-attribute child health dataset. The following configuration was selected based on empirical testing and prior SOM literature:

- **Map Grid Size: 20 × 20 neurons**
  Provides an optimal balance between granularity and computational efficiency for datasets with 10–20 features.
- **Learning Rate: Initialized at 0.1 and linearly decayed to 0.01**
  Ensures fast initial ordering followed by fine-grained convergence.
- **Neighborhood Radius: Varied from 10 to 1 (Gaussian decay)**
  Supports broad neighborhood adaptation early in training, narrowing toward local adjustment.
- **Iterations: 1,000**
  Sufficient for stabilizing the weight vectors based on preliminary convergence testing.
- **Weight Initialization: Random initialization within the normalized range**
  Sufficient for stabilizing the weight vectors based on preliminary convergence testing.

This configuration allows SOM to effectively cluster children into meaningful risk groups prior to supervised classification.

2. Long Short-Term Memory (LSTM) Configuration

Given that the dataset contains cross-sectional child records rather than longitudinal growth sequences, the LSTM was configured to process structured SOM-enhanced feature vectors rather than true time-series sequences. The configuration is as follows:

- Input Representation: Feature vector consisting of 12 normalized attributes + SOM cluster encoding
- LSTM Units: 64
- Dense Layer: 32 neurons, ReLU activation

*name of corresponding author

- Output Layer: Sigmoid for binary classification
- Optimizer: Adam (learning rate = 0.001)
- Loss Function: Binary Cross-Entropy
- Epochs: 150
- Batch Size: 32
- Data Split: 70% training, 15% validation, 15% testing

Although LSTM is typically designed for sequential data, it is applied here as a deep nonlinear classifier benefiting from gated-cell representations, which allow richer feature abstraction compared to shallow models.

3. Computational Platform and Tools

To support transparency and reproducibility, all software environments and hardware specifications used during model development are explicitly documented below:

- Programming Language: Python 3.10
- SOM Library: MiniSom 2.3
- Deep Learning Framework: TensorFlow 2.12 with Keras API
- Hardware Specifications:
  - ✓ CPU: Intel Core i7-12700H
  - ✓ GPU: NVIDIA RTX 3060 (6 GB VRAM)
  - ✓ RAM: 16 GB

GPU acceleration significantly reduced training time and ensured stable performance across k-fold evaluations.

EXPERIMENTAL DESIGN AND EXPECTED RESULTS

This section outlines the experimental design for evaluating the proposed SOM-LSTM model for stunting risk classification, including the data handling, performance metrics, and the anticipated outcomes of the research.

Experimental Design And Expected Results

The experimental design of this study is based on the structured dataset obtained from a maternity hospital in Medan, consisting of 511 cleaned child health records collected between 2021 and 2025. The dataset includes anthropometric measurements (weight, height/length, age, sex) and complementary clinical and sociodemographic attributes such as disease history, parental education, maternal BMI, household income, and immunization completeness. These variables are aligned with international child growth monitoring standards published by the World Health Organization (2006). All records underwent a comprehensive pre-processing pipeline, including duplicate removal, anomaly correction, normalization of numerical variables, and one-hot encoding of categorical attributes, following established practices in medical data preparation (Soni et al., 2022; Novalina et al., in press).

Following pre-processing, the dataset was partitioned into 70% training, 15% validation, and 15% testing subsets, a widely recommended ratio to ensure model generalization and reduce sampling bias (Kohavi, 1995). The entire experimental setup was implemented using Python 3.10, the MiniSom 2.3 library for SOM clustering (Vettigli, 2019), and TensorFlow 2.12 with the Keras API for LSTM model training (Greff et al., 2017). All experiments were executed using an NVIDIA RTX 3060 GPU, which significantly improved computation efficiency during clustering and deep learning training. This rigorous design provides the empirical foundation for the model performance reported in the Results section. (Novalina et al., 2025 (in press); Soni et al., 2022).

Evaluation Metrics

The performance of the developed SOM-LSTM model will be comprehensively evaluated using a suite of standard classification metrics to ensure its robustness and effectiveness in predicting stunting risk. These metrics provide a multifaceted view of the model's performance, considering both its overall correctness and its ability to correctly identify positive and negative cases:

- Accuracy: This metric measures the overall proportion of correctly classified instances (both stunting and non-stunting) out of the total number of instances in the dataset. While it provides a general understanding of the model's performance, accuracy can be misleading in scenarios with imbalanced datasets, where one class significantly outnumbers the other Measures the proportion of correctly classified instances (both stunting and non-stunting) out of the total instances (Ndagijimana et al., 2023).

Eq. (11) — Accuracy

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

- Precision: Precision quantifies the proportion of true positive predictions among all positive predictions made by the model. It indicates the model's exactness or its ability to avoid false positives.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.v10i1.15434
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In the context of stunting prediction, high precision means that when the model predicts stunting, it is highly likely that the child is indeed stunted.

Eq. (12) — Precision

$$\text{Precision} = \frac{TP}{TP + FP} \quad (12)$$

- Recall (Sensitivity): Recall measures the proportion of true positive predictions among all actual positive instances in the dataset. It indicates the model's completeness or its ability to identify all relevant instances. For stunting prediction, high recall is crucial to minimize false negatives, ensuring that as many stunted children as possible are correctly identified for timely intervention.

Eq. (13) — Precision

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

- F1-Score: The F1-Score represents the harmonic mean of precision and recall, providing a balanced measure of the model's performance. It is particularly useful in evaluating models on imbalanced datasets, as it equally considers both false positives and false negatives, offering a more comprehensive assessment than accuracy alone.

Eq. (14) — F1-Score

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

To ensure the generalization capability of the model and mitigate overfitting, the evaluation will incorporate k-Fold Cross-Validation. This method involves partitioning the dataset into k equally sized subsets (folds). The model is iteratively trained on k−1 folds and validated on the remaining fold, with this process repeating k times to ensure each subset serves as a validation set once. The general formula for cross-validation error is given by:

Eq. (15) — Cross-Validation Error

$$\text{Error} = \frac{1}{k} \sum_{i-1}^{k} \text{Error}_i \quad (15)$$

Expected Results

Based on the proposed methodology and the known strengths of hybrid SOM-LSTM architectures, this research anticipates several key outcomes:

Firstly, a significant increase in the accuracy of stunting risk classification is expected. The integrated SOM-LSTM model is hypothesized to achieve higher accuracy compared to conventional classification methods, thereby enhancing the reliability of early risk detection. This improved accuracy will enable more precise identification of children at risk, which is critical for timely and effective intervention (Guo et al., 2024; Ma et al., 2022; Sinaga et al., 2023).

Secondly, the output of the model is expected to facilitate faster and more appropriate medical decision-making. By providing clear and reliable classification results, medical personnel will be better equipped to design proactive and tailored interventions for children identified as being at risk of stunting. This aligns with the vision of transforming clinical maternity hospitals into Smart Hospitals, where clinical decision-making is underpinned by artificial intelligence (Hasdyna et al., 2024).

Thirdly, this research aims to contribute to the digital transformation of maternity hospitals in Medan towards an AI-based hospital. The successful implementation and validation of the SOM-LSTM model will lay a foundational stone for integrating artificial intelligence into clinical decision-making processes within the hospital setting (D.Ruthvik, 2024).

Finally, the project targets an advancement in Technology Readiness Level (TKT) from 2 to 3. Through rigorous laboratory testing and evaluation of the system's performance, the technological readiness will be elevated, providing a solid basis for further development and eventual real-world implementation within healthcare facilities. In terms of dissemination, the research targets include publications in a reputable international journal, such as the Journal of ICT Research and Application (Q4 SJR 0.212 – Impact Factor 3.36), and a national Sinta 2 journal, specifically RESTI (Rekayasa Sistem dan Teknologi Informasi). Additionally, the registration of Intellectual Property Rights (HKI) for the developed SOM-LSTM-based classification algorithm is a targeted output (Begashaw et al., 2025).

RESULT

This section presents the empirical findings of the SOM–LSTM hybrid model using 511 pre-processed child health

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.v10i1.15434
e-ISSN : 2541-2019
p-ISSN : 2541-044X

records obtained from a maternity hospital in Medan. Performance evaluation was conducted through four analyses: classification metrics, confusion matrix, training curves, and k-fold cross-validation.

Model Performance

The performance of the hybrid model was evaluated using accuracy, precision, recall, and F1-score. The SOM–LSTM model demonstrated superior performance compared to the LSTM-only and Random Forest baselines, as shown in Table 1.

Table 1. Model Performance Comparison

Metric	SOM–LSTM	LSTM Only	Random Forest
Accuracy	0.91	0.86	0.83
Precision	0.89	0.83	0.81
Recall	0.92	0.85	0.79
F1-Score	0.90	0.84	0.80

These results support the hypothesis that SOM-based clustering enhances the structure of input representations, enabling the LSTM model to learn temporal patterns more effectively.

Confusion Matrix

The confusion matrix provides deeper insight into the classification behavior of the hybrid model by showing the distribution of true and false predictions.

- True Positive (TP): 45 — correctly identified stunting cases
- True Negative (TN): 28 — correctly identified non-stunting cases
- False Positive (FP): 4 — non-stunting cases misclassified as stunting
- False Negative (FN): 3 — stunting cases missed by the model

The low number of false negatives (FN = 3) reflects high sensitivity, which is crucial for clinical applications to avoid overlooking children who are at risk.

Training Curves

Two training curves were generated to assess learning stability:

1. **Training vs Validation Accuracy Curve**
   Both curves showed consistent improvement, with minimal divergence, indicating an absence of overfitting.
2. **Training vs Validation Loss Curve**
   Loss curves decreased steadily for both training and validation sets, confirming stable convergence of the model.

Two figures should be inserted here:

- Figure X: Train–Validation Accuracy Curve
- Figure Y: Train–Validation Loss Curve

The curves indicate that the model converges smoothly without signs of overfitting, as training and validation curves progress consistently throughout the epochs.

K-Fold Cross-Validation (k = 5)

To assess the model's generalization capability, 5-fold cross-validation was conducted. The statistical summary is as follows:

- Mean Accuracy: 0.894
- Standard Deviation: 0.021

The small standard deviation indicates that model performance is stable across multiple data subsets.

DISCUSSION

This section discusses the implications of the experimental results, their alignment with the research objectives, and connections to prior literature.

A. Relationship Between Methodology and Results

The integration of SOM and LSTM was designed to combine topological clustering with temporal sequence learning. The results confirm that this hybridization produces measurable improvements:

- **Accuracy increased by 5–8%** compared to LSTM-only.
- **Recall increased by 7–10%,** demonstrating improved sensitivity in detecting true stunting cases.
- The reduction in false negatives (FN = 3) is clinically meaningful, highlighting the model's potential for early detection in hospital settings.

These improvements validate the methodological choice to incorporate SOM as a pre-processing stage before

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

LSTM prediction.

B. Comparison with Previous Studies

The results are consistent with findings in earlier hybrid neural network studies (Guo et al., 2024; Ma et al., 2022; Waqas & Humphries, 2024), which reported that:

- Unsupervised clustering can enhance feature quality.
- LSTM models benefit from structured inputs, especially for health-related time series.

Unlike prior works that used synthetic or public datasets, this study utilizes real clinical data from an Indonesian maternity hospital, contributing novel evidence to the literature.

C. Clinical and Operational Implications

From a practical perspective, the model provides strong advantages:

1. High recall ensures stunted children are identified early, enabling timely nutritional or medical interventions.
2. Low false-positive rate reduces unnecessary follow-ups, optimizing hospital resources.
3. The system aligns with the digital transformation objectives of maternity hospitals aiming to implement AI-assisted decision support systems.

This supports the hospital's strategic vision for developing an AI-driven Smart Hospital ecosystem.

D. Strengths and Limitations

Strengths

- Uses real-world hospital data (511 records).
- Employs a hybrid architecture optimized for both clustering and prediction.
- Delivers strong empirical results validated via cross-validation.

Limitations

- Dataset comes from a single hospital, potentially limiting generalizability.
- Some sociodemographic variables may contain noise or inconsistencies.
- The model currently supports binary classification; further granularity (e.g., mild stunting, severe stunting) may be explored.

E. Path Forward for Future Research

Future work may include:

- Expanding the dataset to multi-hospital collaborations.
- Incorporating attention mechanisms or transformer-based architectures.
- Developing a real-time clinical dashboard for hospital use.
- Extending the model to multi-class or regression-based nutritional risk assessment.

CONCLUSION

This study demonstrates that the hybrid SOM–LSTM architecture provides a robust and empirically validated framework for childhood stunting risk classification using real clinical data from a maternity hospital in Medan. The model achieved strong performance across all evaluation metrics, including an accuracy of 0.91, precision of 0.89, recall of 0.92, and an F1-score of 0.90, outperforming both the LSTM-only and Random Forest baselines. The high recall rate is particularly important in clinical settings, as it minimizes false-negative cases and ensures that children at risk of stunting are accurately identified for timely intervention.

The results confirm that SOM-based clustering enhances the quality of feature representation by capturing latent health and nutritional patterns before temporal modeling in LSTM. This combined capability allows the hybrid model to effectively learn both structural and sequential relationships in the data, leading to measurable performance gains. The findings align with the research objective of supporting early stunting detection and contribute to the broader agenda of integrating artificial intelligence into maternal and pediatric healthcare services.

Beyond its predictive accuracy, this research provides operational value for maternity hospitals seeking AI-driven decision support systems. The model offers a feasible and data-driven approach to strengthen clinical screening, optimize resource allocation, and support digital transformation initiatives. Nonetheless, the study is limited by its single-hospital dataset and binary classification approach.

Future research should extend the dataset to multiple healthcare facilities, incorporate additional socioeconomic and clinical variables, and explore more advanced deep learning models such as attention-based or transformer architectures. Integrating the model into a real-time clinical dashboard and evaluating its usability in hospital workflows also represent promising directions for deployment and scale-up.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

REFERENCES

Bappenas. (2019). *RENCANA PEMBANGUNAN JANGKA MENENGAH NASIONAL 2020-2024*. 315.

Begashaw, G. B., Zewotir, T., & Fenta, H. M. (2025). A deep learning approach for classifying and predicting children's nutritional status in Ethiopia using LSTM-FC neural networks. *BioData Mining*, *18*(1), 11. https://doi.org/10.1186/s13040-025-00425-0

D.Ruthvik, Dr. N. R. (2024). AdvanceDeeplearningformalnutritionassessment. *International Journal of Information Technology and Computer Engineering, 12*, 42-59.

Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

Griffin, J. M., Mathew, J., Gasparics, A., Vértesy, G., Uytdenhouwen, I., Chaouadi, R., & Fitzpatrick, M. E. (2022). Machine-Learning Approach to Determine Surface Quality on a Reactor Pressure Vessel (RPV) Steel. *Applied Sciences*, *12*(8), 3721. https://doi.org/10.3390/app12083721

Guo, F., Li, S., Zhao, G., Hu, H., Zhang, Z., Yue, S., Zhang, H., & Xu, Y. (2024). A SOM-LSTM combined model for groundwater level prediction in karst critical zone aquifers considering connectivity characteristics. *Environmental Earth Sciences*, *83*(9), 267. https://doi.org/10.1007/s12665-024-11567-5

Han, N., Qiao, S., Yuan, G., Huang, P., Liu, D., & Yue, K. (2019). A novel Chinese herbal medicine clustering algorithm via artificial bee colony optimization. *Artificial Intelligence in Medicine*, *101*, 101760. https://doi.org/10.1016/j.artmed.2019.101760

Hasdyna, N., Dinata, R. K., Rahmi, & Fajri, T. I. (2024). Hybrid Machine Learning for Stunting Prevalence: A Novel Comprehensive Approach to Its Classification, Prediction, and Clustering Optimization in Aceh, Indonesia. *Informatics*, *11*(4), 89. https://doi.org/10.3390/informatics11040089

Kementerian Kesehatan Republik Indonesia. (2023). *Laporan Hasil Survei Kesehatan Indonesia (SKI) 2023.*

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, *37*, 52–65. https://doi.org/10.1016/j.neunet.2012.09.018

Ma, Y., Yang, J., Feng, J., Wang, H., Li, Y., & Li, Y. (2022). Load data recovery method based on SOM-LSTM neural network. *Energy Reports*, *8*, 129–136. https://doi.org/10.1016/j.egyr.2021.11.070

Ndagijimana, S., Kabano, I. H., Masabo, E., & Ntaganda, J. M. (2023). Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning Techniques. *Journal of Preventive Medicine and Public Health*, *56*(1), 41–49. https://doi.org/10.3961/jpmph.22.388

Novalina, N., Aksar Tarigan, I. A., Kayla Kameela, F., & Rizkinia, M. (2025). Benchmarking machine learning algorithm for stunting risk prediction in Indonesia. *Bulletin of Electrical Engineering and Informatics*, *14*(3), 2252–2263. https://doi.org/10.11591/eei.v14i3.8997

Sekretariat Wakil Presiden Republik Indonesia. (2021). *Strategi Nasional Percepatan Pencegahan Anak Kerdil (Stunting)*. https://stunting.go.id/

Sinaga, F. M., Pipin, S. J., Winardi, S., Tarigan, K. M., & Brahmana, A. P. (2023a). Analyzing Sentiment with Self-Organizing Map and Long Short-Term Memory Algorithms. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *23*(1), 131–142. https://doi.org/10.30812/matrik.v23i1.3332

Sinaga, F. M., Pipin, S. J., Winardi, S., Tarigan, K. M., & Brahmana, A. P. (2023b). Analyzing Sentiment with Self-Organizing Map and Long Short-Term Memory Algorithms. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *23*(1), 131–142. https://doi.org/10.30812/matrik.v23i1.3332

Soni, A., Fahey, N., Ash, A., Bhutta, Z., Li, W., Simas, T. M., Nimbalkar, S., & Allison, J. (2022). Predictive algorithm to stratify newborns at-risk for child undernutrition in India: Secondary analysis of the National Family Health Survey-4. *Journal of Global Health*, *12*, 04040. https://doi.org/10.7189/jogh.12.04040

Sugihartono Tri, Wijaya Benny, Marini, Alkayes Ahmad Faqih, A. H. A. (2025). Optimizing Stunting Detection through SMOTE and Machine Learning: a Comparative Study of XGBoost, Random Forest, SVM, and k-NN. *Journal of Applied Data Sciences*, *6*, 667–682.

Surucu, O., Gadsden, S. A., & Yawney, J. (2023). Condition Monitoring using Machine Learning: A Review of Theory, Applications, and Recent Advances. *Expert Systems with Applications*, *221*, 119738. https://doi.org/10.1016/j.eswa.2023.119738

Tang, Y., Yu, F., Pedrycz, W., Yang, X., Wang, J., & Liu, S. (2022). Building Trend Fuzzy Granulation-Based LSTM Recurrent Neural Network for Long-Term Time-Series Forecasting. *IEEE Transactions on Fuzzy Systems*, *30*(6), 1599–1613. https://doi.org/10.1109/TFUZZ.2021.3062723

Waqas, M., & Humphries, U. W. (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*, *13*, 102946. https://doi.org/10.1016/j.mex.2024.102946

Wei, J., Zhuo, L., Pan, S., Lian, X., Yao, X., & Fu, X. (2023). HeadTailTransfer: An efficient sampling method to improve the performance of graph neural network method in predicting sparse ncRNA–protein interactions. *Computers in Biology and Medicine*, *157*, 106783. https://doi.org/10.1016/j.compbiomed.2023.106783

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume xx, Number xx, month Year
DOI : https://doi.org/10.33395/sinkron.v10i1.15434
e-ISSN : 2541-2019
p-ISSN : 2541-044X

World Health Organization. (2021). *Levels and trends in child malnutrition: UNICEF/WHO/The World Bank Group joint child malnutrition estimates: key findings of the 2021 edition*.
World Health Organization. (2025). *SDG Target 2.2: End all forms of malnutrition.* The Global Health Observatory. https://www.who.int/data/gho/data/themes/topics/sdg-target-2_2-malnutrition

*name of corresponding author