

AI-StuntDetect: Development of Early Stunting Detection System for Young Children Based on CNN and MediaPipe Pose

Asmaul Husna RS[✉], Purnamawati[✉], Hendra Jaya[✉], and Anas Arfandi[✉]

Doctoral Study Program in Engineering Vocational Education, Makassar State University, South Sulawesi, Indonesia

Corresponding author: Purnamawati (e-mail: purnamawati@unm.ac.id), **Author(s) Email:** Asmaul Husna RS (e-mail: asmaul.husna@student.unm.ac.id), Hendra Jaya (e-mail: hendra.jaya@unm.ac.id), Anas Arfandi (e-mail: anas.arfandi@unm.ac.id)

Abstract Stunting affects a substantial proportion of Indonesian children under five and is monitored through routine anthropometric measurement at posyandu, community health posts staffed largely by volunteer cadres. This measurement depends on calibrated instruments and consistent technique, which are not uniformly available across sites, so a low-cost screening aid could extend coverage where conventional practice is weakest. This study aims to develop AI-StuntDetect, a web-based system that estimates children's height and weight from a single mobile-phone photograph, classifies nutritional status, and evaluates whether its image-derived estimates add value beyond simple demographic predictors. A Convolutional Neural Network estimates weight directly from the image, while MediaPipe Pose landmarks are converted into body-proportion features and regressed onto height with a Random Forest. The models were developed on 117 children (1,593 images) using child-level partitioning and validated externally on 48 children from three independent posyandu. Performance was compared against a mean-value baseline and against age-only and age-plus-sex regressions fitted to the training data, and predicted heights were classified against the WHO and Kemenkes reference standards. Height estimation achieved an MAE of 4.092 cm and an R^2 of 0.863, reducing error by 31.9% relative to the strongest training-derived baseline. Weight estimation achieved an MAE of 1.771 kg and an R^2 of 0.590 but did not improve on the age-only regression (1.675 kg). Nutritional status was classified with 70.8% accuracy and a macro-averaged F1 of 0.546, with weak sensitivity for the intermediate stunted category (0.167). A single photograph therefore supports useful height estimation, whereas image-based weight estimation adds nothing beyond what age predicts. AI-StuntDetect is best characterized as an experimental screening prototype suited to triage under manual confirmation.

Keywords Anthropometry; CNN; MediaPipe Pose; Posyandu; Stunting.

1. Introduction

Malnutrition and overnutrition are conditions in which children experience an imbalance in nutrition, which can be caused by various biological and social factors [1], [2], [3], [4], [5]. The World Health Organization (WHO) states that malnutrition has now become a global health crisis, leading to the death of children under five years old [6], [7], [8], [9]. This situation highlights the need for early intervention to minimize the occurrence of both excess and deficiency in nutrition [10], [11]. Malnutrition can be categorized into three types: undernutrition, micronutrient-related malnutrition, and overweight [12], [13], [14]. In some developing countries, this malnutrition becomes a serious problem [15]. This is due to poor economic conditions and limited knowledge about caring for

preschool-aged children, which naturally makes them vulnerable to various diseases that can lead to death [16], [17], [18]. Indonesia also has one of the highest malnutrition rates. In 2023, the Indonesian Health Survey (SKI) released a report indicating that the prevalence of stunting was 21.5% among children under 5 years old, with wasting at 8.5% and underweight at 15.9%. Although this is still far from the national target of 14%, the numbers have been decreasing year by year [19], [20].

Furthermore, in South Sulawesi Province (Sulsel), the stunting prevalence in 2023 was 27.4%, far from the national target and above the national SKI figure, and the province has consistently ranked among those with the highest stunting prevalence in Indonesia [21], [22]. Of course, this condition requires special attention

to reduce stunting effectively and efficiently in the Susel area. One approach is to assist health workers in conducting contactless screening using Artificial Intelligence (AI) technology, whether through Machine Learning (ML) or Deep Learning (DL).

Several studies related to the use of DL for anthropometric measurements have been conducted. Ichikawa et al. [23] estimated body weight from chest and abdominal scout CT scan images using CNN algorithms. They achieved a Mean Absolute Error (MAE) of 2.75 kg for chest datasets and 4.77 kg for abdominal datasets. A similar study was conducted by Demircioğlu et al. [24], who used CT localizers of pediatric and young adult patients with EfficientNetV2-S, ResNet-18, and ResNet-34 to estimate height and weight. This research achieved an MAE of 5.58 ± 4.26 cm for height and 4.25 ± 4.28 kg for weight. Another study [25] also used a CNN to estimate height and weight from clinical data obtained at various angles, including posterior, anterior, and two lateral views. This study resulted in an MAE of 3.7 kg for weight and 3.6 cm for height. Not only studies using 2D data, but other research such as [26] used 3D data to measure weight in Emergency Care Settings. This study also employed a CNN pipeline and achieved an MAE of 5.4%. Additionally, [27] collected 200 newborn photos to train models using MobileNet, VGG16, GoogleNet, and AlexNet, resulting in average MAEs of 1.1 cm and 0.28 kg.

There is also research that uses Yolo to detect human body joint points and combines it with ML to perform height regression, as done by [28], with the AdaBoost Regressor achieving an MAE of 6.2 cm. Studies [29] and [30] use MediaPipe Pose to detect children's joints, measure height using a doll dataset, and test on toddlers. Study [31] measures adult height using YOLOv8 and the MediaPipe Library combined with multivariate linear regression, achieving an average error of 1.2%.

Despite this progress, four gaps remain in the literature reviewed above. First, the populations studied are poorly matched to community stunting screening: the CT-based approaches were developed on hospital patients whose ages extend well beyond early childhood, adults in [23] and patients up to 21 years in [24], the 3D imaging study [26] on emergency-care patients, [27] on newborns, and [29], [30] on dolls, so evidence on real children aged 0 to 59 months photographed during routine growth monitoring remains limited. Second, most studies report performance on a held-out portion of a single dataset, and validation on children, sites, and operators independent of model development is rarely presented. Third, errors are typically reported as absolute metrics without a baseline, leaving it unclear whether image-derived estimates add anything beyond what age and

sex already predict; this matters particularly in young children, whose narrow anthropometric ranges can make a trivial predictor appear accurate. Fourth, anthropometric estimates are seldom translated into nutritional status categories, so the accuracy with which such systems classify children as normal, stunted, or severely stunted, which is the decision that screening actually supports, has rarely been reported.

To address these gaps, this study proposes AI-StuntDetect, a hybrid pipeline that estimates anthropometric measures from a single RGB photograph. A Convolutional Neural Network estimates body weight directly from the image, while MediaPipe Pose landmarks are converted into distance and body-proportion features and regressed onto height using a Random Forest Regressor. The predicted height is then converted into a height-for-age Z-score and classified against the WHO and Kemenkes reference standards. Both models are served through a FastAPI endpoint and integrated into a Laravel web application, so that posyandu cadres can obtain the estimates and a nutritional status classification without installing any software.

The objective of this study is threefold: (i) to develop a low-cost, web-based system that enables contactless stunting screening at posyandu level from a single mobile-phone photograph; (ii) to evaluate its height and weight estimates on an external cohort from independent posyandu against baselines realizable at deployment time, rather than through absolute error metrics alone; and (iii) to determine how accurately the resulting estimates translate into WHO and Kemenkes nutritional status categories. The individual components used here, namely convolutional networks, MediaPipe Pose, Random Forest regression, and the FastAPI and Laravel frameworks, are established technologies, and we claim no novelty in their combination. The contributions of this study are as follows: (1) a child-level data partitioning protocol that prevents near-duplicate images of the same child, several of which were captured per session, from appearing in both training and evaluation subsets; (2) an external validation at three independent posyandu with children, operators, lighting, and clothing not involved in model development; (3) a pose-derived feature set that combines inter-landmark distances with body-proportion ratios; and (4) a baseline-referenced evaluation against age-only and age-plus-sex regressions, together with a stunting classification analysis, which shows that image-derived information improves substantially on demographic prediction for height but not for weight.

The paper is structured as follows: Section II describes the datasets and methods, including the CNN-based weight estimation model and the MediaPipe Pose–Random Forest Regressor approach

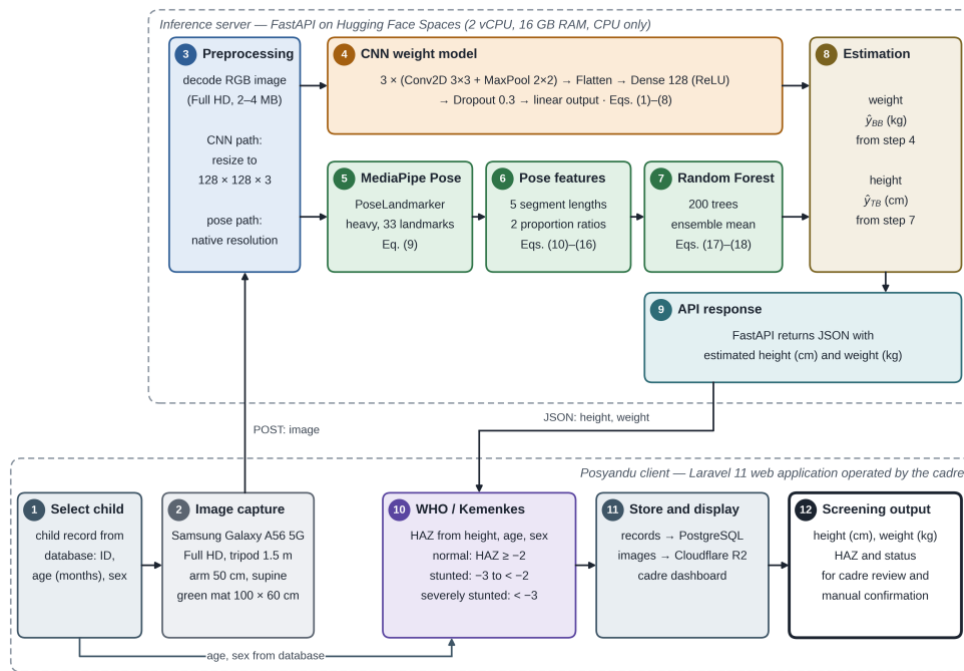


Fig. 2. System-level architecture of AI-StuntDetect, from image acquisition at the posyandu to the final screening output. Steps 1–2 and 10–12 run in the Laravel web application at the posyandu; steps 3–9 run on the FastAPI inference server. Equation numbers refer to the formulations in Section II.

for height estimation. Section III presents the results of the experiment, including external validation

performance, comparison against deployment-realizable baselines, and stunting classification against the WHO/Kemenkes reference standard. Section IV discusses the findings, clinical implications for stunting screening at the posyandu level, and limitations of the study. Finally, Section V concludes the paper and suggests directions for future research.

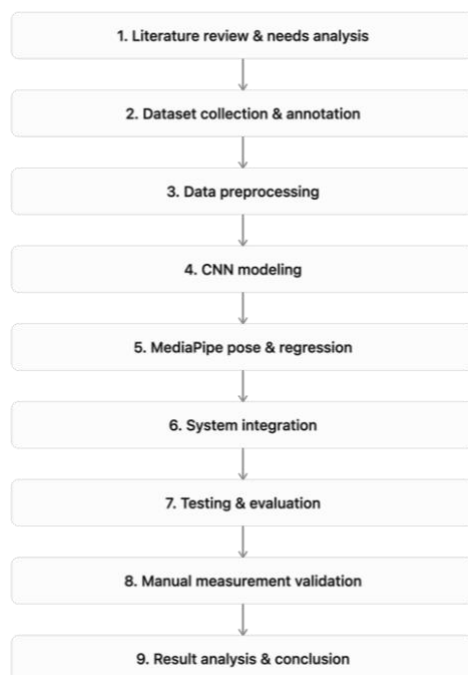


Fig. 1. Flowchart of the proposed methodology, spanning literature review, dataset preparation, modeling, and manual measurement validation.

II. Method

A. Types and Approaches of Research

This research is an applied study that the community, especially posyandu officers, can use to screen preschool children. A quantitative approach was chosen to continue the research and development process. The main goal of this study is to develop an AI-based web-based early detection system for stunting in young children. Integrating AI into the website makes its use more flexible, as health workers do not need to install an application [32]. The overall research workflow is shown in Fig. 1, the system architecture in Fig. 2, and the complete inference procedure in Algorithm 1; each is described in the subsections that follow. This study uses two AI approaches. First, an image-based CNN is used to determine children's body weight. Second, the use of MediaPipe Pose to generate the child's body coordinates and translate them into height estimation assisted by Random Forest. Then, after obtaining height and weight, the next step is to perform anthropometric indicator calculations as specified by

Algorithm 1. Pseudocode of the AI-StuntDetect inference procedure, covering image input, preprocessing, CNN weight prediction, landmark extraction, pose-feature computation, Random Forest height prediction, Z-score calculation, and nutritional status determination. Lines (1)–(2) and (19)–(27) run in the Laravel web application; lines (3)–(18) run on the FastAPI inference server.

Input: child record c with age a (months) and sex s ; RGB image I captured under the acquisition protocol
Output: estimated weight $\hat{y}_{BB}$ (kg), estimated height $\hat{y}_{TB}$ (cm), height-for-age Z-score HAZ, nutritional status C

Client (Laravel web application)

- (1) select child record c ; read a and s from the database
- (2) capture image I ; send I to the FastAPI endpoint

Inference server (FastAPI)

- (3) decode I as an RGB array
- (4) $I_{cnn} \leftarrow \text{Resize}(I, 128 \times 128 \times 3)$
- (5) $\hat{y}_{BB} \leftarrow \text{CNN}(I_{cnn}) \triangleright \text{Eqs. (1)–(8)}$
- (6) $\mathcal{L} \leftarrow \text{PoseLandmarker}(I, \text{num_poses} = 1) \triangleright 33 \text{ landmarks}$
- (7) **if** $\mathcal{L} = \emptyset$ **then**
- (8) **return** error “no pose detected” $\triangleright$ image rejected
- (9) **end if**
- (10) **for** $j \in \{\text{shoulder, hip, knee, ankle}\}$ **do**
- (11) $P_j \leftarrow (P_j^{\text{left}} + P_j^{\text{right}}) / 2 \triangleright \text{mean of bilateral landmarks}$
- (12) **end for**
- (13) $P_{\text{head}} \leftarrow \text{landmark among } \mathcal{L}_0, \dots, \mathcal{L}_{10} \text{ with minimum } y$
- (14) compute torso_length, upper_leg_length, lower_leg_length, leg_length, total_visible_height $\triangleright \text{Eqs. (9)–(14)}$
- (15) compute torso_leg_ratio, upper_lower_leg_ratio $\triangleright \text{Eqs. (15)–(16)}$
- (16) $x \leftarrow \text{vector of the seven features from lines (14)–(15)}$
- (17) $\hat{y}_{TB} \leftarrow (1/B) \sum_{b=1..B} T_b(x), B = 200 \triangleright \text{Eq. (17)}$
- (18) **return** JSON $\{\hat{y}_{TB}, \hat{y}_{BB}\}$ to the client

Client (Laravel web application)

- (19) $(M, S) \leftarrow \text{reference median and coefficient of variation for } (s, a) \triangleright \text{WHO / Kemenkes table}$
- (20) $\text{HAZ} \leftarrow (\hat{y}_{TB} / M - 1) / S$
- (21) **if** $\text{HAZ} < -3$ **then** $C \leftarrow \text{“severely stunted”}$
- (22) **else if** $\text{HAZ} < -2$ **then** $C \leftarrow \text{“stunted”}$
- (23) **else** $C \leftarrow \text{“normal”}$
- (24) **end if**
- (25) store $(c, \hat{y}_{TB}, \hat{y}_{BB}, \text{HAZ}, C)$ in PostgreSQL and I in Cloudflare R2
- (26) display $\hat{y}_{TB}, \hat{y}_{BB}, \text{HAZ}$ and C to the cadre for review
- (27) **return** $\hat{y}_{BB}, \hat{y}_{TB}, \text{HAZ}, C$

the Indonesian government through Minister of Health Regulation No. 20 of 2020 via the website. The stunting categories used are normal, stunted, or severely stunted [33].

B. Research Framework

Fig. 1 shows the workflow of this study, starting with the literature review and needs analysis [34]. Then, it continues with data collection from posyandu conducting routine activities. After the data is collected, the next step is data preprocessing to standardize image resolution and size. Once the data is ready, the

next step is to build a CNN model for weight estimation and use MediaPipe Pose with regression for height estimation. The resulting model is then integrated into a website. The website is built using Laravel, and FastAPI is used to embed the model; the model is deployed on Hugging Face. After successful integration, the system is tested with test data and through live testing at the posyandu to observe its real-time performance. The final stage is to discuss the results and draw conclusions based on the research conducted.

Fig. 2 presents the system-level architecture of AI-StuntDetect, which is divided between a posyandu client and an inference server. In the Laravel web application, the cadre first selects the child to be measured, whose identifier, age in months, and sex are held in the application database (block 1), and then captures the image under the acquisition protocol described in the Data Collecting subsection (block 2). Only the image is sent to a FastAPI endpoint hosted on Hugging Face Spaces, where it is decoded and prepared along two paths (block 3): a copy resized to $128 \times 128 \times 3$ pixels is passed to the CNN weight model (block 4), while the image at native resolution is passed to MediaPipe Pose (block 5). The detected landmarks are converted into segment lengths and body-proportion ratios (block 6), from which the Random Forest regressor predicts height (block 7). The two estimates (block 8) are returned to the client as a JSON response (block 9). The Laravel application then combines the estimated height with the child's age and sex to compute the height-for-age Z-score and assign the nutritional status category against the WHO and Kemenkes thresholds (block 10), stores the record in PostgreSQL and the image in Cloudflare R2 (block 11), and displays the result to the cadre for review and manual confirmation (block 12). The equations referenced in each block are detailed in the subsections that follow.

The complete inference procedure is formalized in Algorithm 1. Lines 1–2 run in the Laravel application, where the cadre selects the child record and captures the image. Lines 3–18 run on the inference server: the image is resized for the CNN weight model, pose landmarks are extracted and screened for detection failure, the seven pose-derived features are computed, and the Random Forest predicts height. Lines 19–27 run in the Laravel application again, where the estimated height is combined with the child's age and sex to compute the height-for-age Z-score against the WHO and Kemenkes reference, the nutritional status category is assigned, and the result is stored and displayed for the cadre's review.

C. Subject and Object of the Research

The research was conducted in Jeneponto Regency, South Sulawesi Province, focusing on children aged 0–59 months. The data to be collected in this study are image data of children. Data collection from children requires informed consent, which will be explained beforehand to the parents or guardians of the children whose data will be collected.

D. Research Tools and Materials

Fig. 3 shows the hardware used in this study, consisting of a tripod with a height of 1.5 meters that has an overhead arm, a smartphone with a minimum camera resolution of 1080 pixels, and a green mat

measuring 1×0.6 meters. Additionally, a scale and a measuring tape will be used as ground truth for testing. A MacBook Air M2 laptop is used for model training process. A single capture device, a Samsung Galaxy A56 5G smartphone operating at Full HD resolution, was used for every image in both the development and external validation phases, so that sensor characteristics, lens geometry, and image resolution were constant across the entire dataset.

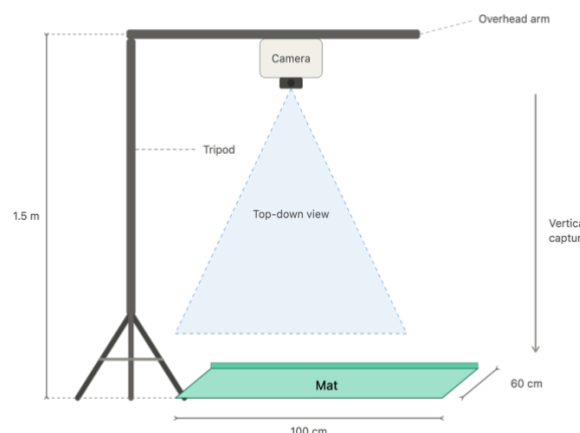


Fig. 3. Camera positioning setup for image acquisition, with the device fixed at 1.5 m height to capture a top-down view of the measurement mat (100 cm × 60 cm).

The software used in this research employs Python 3 and TensorFlow to build and train the CNN architecture. MediaPipe Pose Library and Random Forest are used to detect body coordinates and perform regression results. For training and testing the model, Google Colab Pro is utilized. On the website side, Laravel 11, based on PHP and PostgreSQL as the database are used. We do not embed the model directly into the website, so we use Hugging Face and create FASTAPI to make the model lighter when run by the website in real time.

E. Data Collection

The image data used in this study are primary data collected from 8 posyandu in Jeneponto Regency. The selected posyandu include Bontoa Village, Pabentengan Village, Jenetallasa Village, Lengkelengkesse Village, Bonia Village, Lianga Village, Kunjungmangge Village, and Bontomanai Village. There are 117 children aged between 0 and 59 months. From the image data collection, a total of 1,593 images were obtained. Before collecting data, we provided parents with a consent letter explaining what data would be collected and that it would not be published to protect the children's privacy. Data collection was

Table 1. Controlled and uncontrolled image acquisition variables across both study phases. Capture device, camera geometry, background, and subject posture were held constant, whereas illumination and clothing varied with the conditions of each facility.

Acquisition variable	Development phase	External validation
Capture device	Samsung Galaxy A56 5G	Same device
Image resolution	Full HD	Full HD
Camera height	1.5 m (fixed)	1.5 m (fixed)
Overhead arm	50 cm	50 cm
Mounting	Tripod	Tripod
Background	Green mat, 100 × 60 cm	Green mat, 100 × 60 cm
Subject posture	Supine, facing camera	Supine, facing camera
Illumination	Ambient room light (varied)	Ambient room light (varied)
Clothing	Not standardized	Not standardized

Table 2. Characteristics of the development and external validation cohorts. The two groups did not differ significantly in sex ($p = 0.893$), height ($p = 0.252$), or weight ($p = 0.246$), but the validation cohort was younger on average ($p = 0.032$).

Characteristic	Development (n = 117)	External validation (n = 48)
Boys, n (%)	53 (45.3)	23 (47.9)
Girls, n (%)	64 (54.7)	25 (52.1)
Age, months (mean ± SD)	33.5 ± 17.3	27.4 ± 16.2
Age, months (range)	0 – 60	2 – 67
Height, cm (mean ± SD)	86.4 ± 14.1	83.8 ± 13.3
Height, cm (range)	57.6 – 109.0	56.8 – 107.0
Weight, kg (mean ± SD)	11.39 ± 3.42	10.73 ± 3.27
Weight, kg (range)	5.4 – 20.0	4.7 – 18.2
Nutritional status		37 normal, 6 stunted, 5 severely stunted

carried out using a setup as shown in Fig. 3. The children were laid on a green mat facing the camera, with a distance of 1.5 meters between the camera and the child. Between 12 and 15 images were taken of each child, with the body in continuous movement so the arms and legs were not in the same position in every frame. Images affected by motion blur were excluded during a manual screening step before training. This was done so that the dataset we obtained could demonstrate that the changing positions of children during image capture can help the model maintain its consistency. In addition to image data collection, we also performed manual measurements using a digital scale and a height meter, which served as labels during model training.

Image acquisition followed an identical protocol in the development and external validation phases. The child was placed supine on a green mat measuring 100 × 60 cm, facing the camera, with the camera mounted on a tripod-supported overhead arm extending 50 cm at a fixed height of 1.5 meters. Table 1 summarizes which acquisition variables were controlled and which were allowed to vary. Camera device, camera height, overhead-arm geometry, mat dimensions, and subject posture were held constant in both phases. Two variables were not standardized: illumination relied on the ambient light available in each facility, with no

supplementary lighting, and clothing was not standardized, with children being photographed in the garments they were wearing at the time of the visit.

Reference anthropometry was obtained by posyandu cadres following the routine growth-monitoring procedures of the Indonesian Ministry of Health (Permenkes No. 2/2020). At each session, two cadres performed the measurements, one for weight and one for stature, after a briefing by the research team on the study protocol and capture apparatus. Children below 12 months were weighed on a Metrisis digital infant scale and measured supine with an infantometer; older children were weighed on a GEA standing scale and measured standing with a stadiometer. Weight was recorded to 0.01 kg and stature to 0.1 cm, with all children measured barefoot. Instruments were calibrated before first use and re-checked at the start of each session. Each child was measured once per session, and the values were cross-checked against the child's Maternal and Child Health (KIA) book; none required revision.

Table 2 reports the characteristics of both cohorts. The development cohort comprised 117 children (45.3% boys) with a mean age of 33.5 ± 17.3 months, a mean height of 86.4 ± 14.1 cm and a mean weight of 11.39 ± 3.42 kg. The external cohort comprised 48

children (47.9% boys) with a mean age of 27.4 ± 16.2 months, a mean height of 83.8 ± 13.3 cm and a mean weight of 10.73 ± 3.27 kg; of these, 37 were of normal nutritional status, 6 were stunted and 5 were severely stunted. Sex distribution did not differ significantly between cohorts ($\chi^2 = 0.018$, $p = 0.893$), nor did height ($p = 0.252$) or weight ($p = 0.246$). The external cohort was younger on average ($p = 0.032$), reflecting the age composition of children attending the participating posyandu during the collection period.

F. Preprocessing

In the data preprocessing stage, the collected images are processed through a simple step, which is resizing the images to a fixed dimension of 128×128 pixels. This process aims to standardize the size of all input images before they are processed by the model, since images obtained from acquisition generally have different resolutions and aspect ratios depending on the camera device used. Standardizing the size to 128×128 pixels aligns with the input dimension requirements of the model architecture and helps reduce computational load during training compared to using the original larger image resolutions [35], [36].

G. Estimated Body Weight Model (CNN)

The CNN architecture used in this study is built using a sequential model approach with input images measuring $128 \times 128 \times 3$ pixels. Feature extraction is performed through three convolutional (Conv2D) layers arranged sequentially with increasing numbers of filters, namely 32, 64, and 128 filters with a 3×3 kernel size, each using the ReLU (Rectified Linear Unit) activation function to introduce non-linearity into the model. Each convolutional layer is followed by a 2×2 MaxPooling2D layer, which gradually reduces the spatial dimensions of the feature maps while preserving salient features and reducing computational complexity in subsequent layers. The deeper the convolutional layer, the more complex the features the model can capture from simple edges and textures in the early layers to more abstract body shape patterns in the deeper layers [37].

After feature extraction process through three convolutional blocks is complete, the feature map output is flattened into a one-dimensional vector through a Flatten layer before being passed to the fully connected layer. This part consists of a Dense layer with 128 neurons and a ReLU activation function, followed by a Dropout layer with a ratio of 0.3 to prevent overfitting by randomly deactivating some neurons during training. As the output layer, a Dense layer with one neuron and a linear activation function is used, which is suitable for regression problems because it produces continuous predicted values in this case, an estimate of the child's weight without restricting the output value range as classification activation functions

do. Formally, the convolution operation performed by each Conv2D layer is expressed in Eq. (1),

$$S(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n) \quad (1)$$

with $S(i, j)$ is the output value at position (i, j) of the resulting feature map, I is the input feature map, K is the 3×3 convolution kernel, and m, n are the indices iterated across the kernel window. This equation shows how each output pixel is computed as the weighted sum of the input region under the kernel, allowing the network to detect local spatial patterns such as edges and body contours in the input image [37], [38], [39].

The spatial dimension of the resulting feature map after each convolution is given by Eq. (2),

$$O = \frac{W - F + 2P}{s} + 1 \quad (2)$$

where O is the output dimension, W is the input dimension, F is the kernel size, P is the padding, and s is the stride [38]. This equation accounts for the reduction from 128×128 to 126×126 pixels after the first convolutional layer, since a 3×3 kernel is applied with no padding and a stride of 1. The ReLU activation applied after each convolutional layer is defined in Eq. (3),

$$f(x) = \max(0, x) \quad (3)$$

with $f(x)$ is the activated output and x is the pre-activation input to the unit. This function introduces non-linearity by passing positive values unchanged while suppressing negative values to zero, which prevents the vanishing-gradient problem common in deeper networks [40]. The max-pooling operation that follows each convolutional block is given by Eq. (4),

$$Q(i, j) = \max_{(m, n) \in R_{ij}} I(m, n) \quad (4)$$

with $Q(i, j)$ the pooled output at position (i, j) , $I(m, n)$ the input value at position (m, n) , and R_{ij} the non-overlapping 2×2 input window corresponding to output position (i, j) [39]. This operation reduces the spatial dimension of the feature map while retaining the most salient activation in each region, thereby lowering computational cost in subsequent layers. The flattened feature vector is subsequently passed through the fully connected (Dense) layer, formalized in Eq. (5) [39],

$$z = f\left(\sum_{i=1}^n w_i x_i + b\right) \quad (5)$$

with z is the output of the neuron, w_i is the weight connecting input i , x_i is the i -th input from the previous layer, b is the bias term, n is the number of input units, and f is the ReLU activation function. This equation shows how the 25,088 flattened units are combined into 128 higher-level representations before being passed to the output layer.

Table 3. Training configuration of the weight estimation network. Of the 3,304,769 parameters, only 93,248 reside in the convolutional blocks; the remainder are concentrated in the fully connected layer following flattening.

Setting	Value
Input size	128 × 128 × 3
Optimizer	Adam (Keras defaults: $\text{lr} = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$)
Loss function	Mean squared error
Batch size	32
Maximum epochs	30
Early stopping	monitor = val_loss, patience = 5, restore_best_weights = True
Regularization	Dropout 0.3 (no L1/L2 penalty, no batch normalization)
Data augmentation	None
Total parameters	3,304,769 (all trainable)

To mitigate overfitting, the Dropout layer randomly deactivates a proportion of the activations during training, as expressed in Eq. (6),

$$\hat{h}_i = \frac{(h_i \cdot m_i)}{(1 - p)}, m_i \sim \text{Bernoulli}(1 - p) \quad (6)$$

with $\hat{h}_i$ the i -th activation after dropout, h_i the corresponding activation before dropout, m_i a binary mask value drawn from a Bernoulli distribution with retention probability $1 - p$, and p the dropout rate, set at 0.3 in this study [41]. Division by $(1 - p)$ keeps the expected activation equal between training and inference. This regularization forces the network to avoid over-reliance on any single neuron, improving generalization to unseen data. The final output layer, which produces the predicted body weight, is given by Eq. (7) [38],

$$\hat{y}_{BB} = w_o \cdot h + b_o \quad (7)$$

with $\hat{y}_{BB}$ is the predicted weight, h is the output of the last Dense layer, and w_o , b_o are the weight and bias of the output neuron. A linear activation function is used at this layer to allow an unrestricted continuous output, appropriate for a regression task.

The entire network is trained by minimizing the mean squared error loss function shown in Eq. (8) [38], [42],

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (8)$$

with $\mathcal{L}$ is the loss value, N is the number of samples in a training batch, y_i is the actual weight, and $\hat{y}_i$ is the predicted weight for sample i . This loss function penalizes larger prediction errors more heavily, guiding the optimizer toward parameters that minimize the overall squared deviation between predicted and measured weight.

Table 3 summarizes the training configuration. The network was trained from scratch; no pre-trained backbone or transfer learning was used. The parameter count is large relative to the number of contributing children, and it is worth noting where that

capacity resides. The three convolutional blocks together account for only 93,248 parameters, approximately 2.8% of the total. The remaining 3,211,392 parameters, some 97.2% of the model, are concentrated in the single fully connected layer following flattening, because the $14 \times 14 \times 128$ feature map produces a 25,088-dimensional vector. The effective capacity of the feature extractor is therefore modest, and the parameter count is driven almost entirely by the flatten-to-dense transition rather than by the depth of the convolutional stack. Overfitting was controlled by dropout at rate 0.3 before the output layer and by early stopping on validation loss.

H. Estimated Height Model (MediaPipe Pose)

At this stage, MediaPipe Pose is used to 'read' the child's body position from a photo by detecting key body points such as the nose, shoulders, hips, knees, and ankles. Since the positions of these points are expressed in normalized vertical coordinates (y -axis), the system can calculate distances between them for example, body length (from shoulder to hip), thigh length (from hip to knee), and calf length (from knee to ankle). Formally, the distance between any two landmarks is computed using the Euclidean distance formula shown in Eq. (9) [43], [44],

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (9)$$

with $d(P_1, P_2)$ is the distance between two landmark points, and (x_1, y_1) , (x_2, y_2) are the normalized coordinates of points P_1 and P_2 , respectively. This formula converts the raw pose landmarks detected by MediaPipe into interpretable linear measurements of body segments. Applying Eq. (9) to the shoulder and hip landmarks yields the torso length feature, shown in Eq. (10) [29], [44],

$$\text{torso_length} = d(P_{\text{shoulder}}, P_{\text{hip}}) \quad (10)$$

with P_{shoulder} and P_{hip} are the landmark coordinates of the shoulder and hip, respectively. The upper leg length is computed in Eq. (11) [29],

$$\text{upper_leg_length} = d(P_{\text{hip}}, P_{\text{knee}}) \quad (11)$$

with P_{hip} and P_{knee} are the landmark coordinates of the hip and knee, respectively. Similarly, the lower leg length is computed in Eq. (12) [29],

$$\text{lower_leg_length} = d(P_{knee}, P_{ankle}) \quad (12)$$

with P_{knee} and P_{ankle} are the landmark coordinates of the knee and ankle, respectively. The total leg length used as a predictor feature is obtained by summing the two segments, as shown in Eq. (13) [29],

$$\text{leg_length} = \text{upper_leg_length} + \text{lower_leg_length} \quad (13)$$

with upper_leg_length and lower_leg_length obtained from Eq. (11) and Eq. (12), respectively. The total visible height feature is computed in Eq. (14) [29],

$$\text{total_visible_height} = d(P_{head}, P_{ankle}) \quad (14)$$

with P_{ankle} the ankle landmark coordinate and P_{head} the head reference point. Because the MediaPipe Pose topology defines no landmark at the vertex of the skull, P_{head} is approximated as the most superior detected facial landmark, computed as the minimum vertical coordinate across the eleven head landmarks of the topology (the nose, the six eye landmarks, the two ear landmarks, and the two mouth landmarks). This approximation excludes the height of the calvaria above the facial plane and therefore underestimates true stature by a small, age-varying amount, a consequence examined in the age-stratified error analysis. The two ratio features used in this study are defined as follows. Eq. (15) gives the ratio of torso length to total leg length [29],

$$\text{torso_leg_ratio} = \frac{\text{torso_length}}{\text{leg_length}} \quad (15)$$

with torso_leg_ratio the resulting ratio, and torso_length and leg_length obtained from Eq. (10) and Eq. (13), respectively. Eq. (16) gives the ratio of upper leg length to lower leg length [29],

$$\text{upper_lower_leg_ratio} = \frac{\text{upper_leg_length}}{\text{lower_leg_length}} \quad (16)$$

where $\text{upper_lower_leg_ratio}$ is the resulting ratio, and upper_leg_length and lower_leg_length are obtained from Eq. (11) and Eq. (12), respectively. These ratio features were included on the rationale that body proportions remain comparatively stable when the camera-to-subject distance varies between captures, whereas raw distances do not. Because the acquisition protocol held the camera-to-subject distance constant throughout both study phases, however, this property could not be tested empirically here, and it therefore remains a design rationale rather than a validated result.

After all pose-derived features have been extracted, they are used to train a Random Forest Regressor, whose hyperparameters are listed in Table 4, and which combines the predictions of many decision trees through the ensemble averaging process formally expressed in Eq. (17) [45],

$$\hat{y}_{TB} = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (17)$$

with $\hat{y}_{TB}$ is the final predicted height, B is the number of decision trees in the forest, and $T_b(x)$ is the prediction of the b -th tree given the feature vector x . This averaging process reduces the variance associated with any single decision tree, improving the stability of the height prediction.

One of the advantages of Random Forest is its ability to quantify the contribution of each feature to the overall prediction, computed as the mean decrease in impurity across all trees, as shown in Eq. (18) [46],

$$I(x_j) = \frac{1}{B} \sum_{b=1}^B \sum_{t \in T_b} p(t) \Delta i(t, x_j) \quad (18)$$

with $I(x_j)$ is the importance score of feature x_j , B is the number of trees, $p(t)$ is the proportion of samples reaching node t , and $\Delta i(t, x_j)$ is the decrease in impurity achieved by splitting node t on feature x_j . The resulting importance scores for each pose-derived feature are reported in Section III.D.

Table 4. Random Forest hyperparameter search space and the configuration selected by exhaustive grid search with five-fold cross-validation, optimizing the negative mean absolute error across 36 candidate configurations and 180 model fits.

Parameter	Search space	Selected value
n_estimators	100, 200, 300	200
max_depth	None, 10, 20	None
min_samples_split	2, 5	2
min_samples_leaf	1, 2	1
max_features	not tuned	scikit-learn default
random_state	fixed	42

Pose estimation was performed with MediaPipe 1.0.1 using the Tasks API (PoseLandmarker), which supersedes the legacy mp.solutions interface. The heavy variant of the pose landmarker model (pose_landmarker_heavy, float16 quantization) was used throughout, configured with num_poses = 1 and operating in image running mode. The remaining detection parameters retained their library defaults of 0.5 for minimum pose detection confidence, minimum pose presence confidence, and minimum tracking confidence. No per-landmark visibility threshold was applied: all landmarks returned for a successfully detected pose were used in feature extraction. Images in which no pose was detected were excluded from the dataset, with no interpolation or imputation of missing

landmarks. For structures represented by bilateral landmarks (shoulders, hips, knees, and ankles), the vertical coordinate was taken as the arithmetic mean of the left and right values.

Random Forest hyperparameters were selected by exhaustive grid search with five-fold cross-validation (KFold with shuffling, random_state = 42), optimizing the negative mean absolute error over 36 candidate configurations and 180 model fits. Table 4 reports the search space and the selected configuration: 200 trees with no depth limit, a minimum of two samples to split a node and one sample per leaf. The search therefore favoured fully grown trees over the more strongly regularized alternatives in the grid. No data augmentation was applied, since geometric transformations such as rotation, scaling, or shearing would alter the pixel-space body proportions from which stature is inferred and would corrupt the relationship between the extracted features and the ground-truth measurement.

I. System Integration

To perform real-time model integration on a web-based platform, we use two separate architectures. The model obtained during training is deployed on Hugging Face, which serves as the backend engine using FastAPI. Next, FastAPI validates the input and output in JSON structure. Then, a Laravel-based web application acts as a client-oriented application that sends asynchronous HTTP requests to the FastAPI endpoint and efficiently presents the prediction results to the end user. Measurement data will be stored in PostgreSQL, and measurement image data will be stored in Cloudflare R2.

J. System Testing

The predictive performance of the image-based regression model in estimating anthropometric indicators is evaluated using three standard statistical metrics: MAE, MSE, and RMSE. MAE measures the average magnitude of absolute errors without considering the direction of deviation, as formalized in Eq. (19) [42],

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (19)$$

with n is the number of samples, y_i is the actual measured value, and $\hat{y}_i$ is the predicted value for sample i . MSE, shown in Eq. (20) [42], measures the average squared difference between the estimated and manually measured values, thereby placing a greater penalty on large prediction errors, Eq. (20) with n , y_i and $\hat{y}_i$ as defined for Eq. (19).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (20)$$

RMSE, given in Eq. (21) [42], is the square root of MSE and returns the error scale to its original measurement unit (centimeters or kilograms), making the magnitude of deviation easier to interpret. The combination of these three metrics describes the model's precision, stability, and error tolerance.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (21)$$

with MSE as defined in Eq. (20). To provide a reference point for interpreting these metrics, a constant mean-value baseline was constructed for each target variable, formally defined in Eq. (22) [42],

$$\hat{y}_{baseline} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (22)$$

where $\hat{y}_{baseline}$ is the baseline prediction, $\bar{y}$ is the arithmetic mean of the observed values, n is the number of samples in the test set, and y_i is the actual value of sample i . For every sample in the external test set, the baseline predicts this mean value (83.746 cm for height and 10.727 kg for weight), irrespective of the input image. Model performance is then expressed as the percentage reduction in error relative to this baseline, computed using Eq. (23) [42],

$$Improvement(\%) = \frac{E_{baseline} - E_{model}}{E_{baseline}} \times 100\% \quad (23)$$

where $E_{baseline}$ and E_{model} are the error metric (MAE, MSE, or RMSE) of the baseline and the proposed model, respectively. A positive value indicates that the model outperforms the baseline, while a negative value indicates that the model performs worse than simply predicting the mean.

This comparison is necessary because absolute error magnitudes are not comparable across variables measured in different units and spanning different natural ranges; a model that appears accurate in absolute terms may perform no better than a trivial predictor. A mean-value baseline alone, however, is not a demanding test. We therefore additionally evaluate three baselines estimated exclusively from the training subset and applied unchanged to the external cohort: the training-set mean, an age-only linear regression, and an age-plus-sex linear regression. Unlike a baseline derived from the evaluation cohort, these are realizable at deployment time and establish whether image-derived information contributes beyond what demographic data alone provide.

The dataset of 117 children (1,593 images) was stratified by age group (six bins: 0–6, 7–12, 13–24, 25–36, 37–48, and 49–60 months) and sex, and split at the child level into training, validation, and internal test subsets using a 70:15:15 ratio applied within each stratum. This allocated 84 children (1,115 images) to

training, 19 children (239 images) to validation, and 14 children (239 images) to the internal test set, ensuring that no child, nor any of their associated images, appeared in more than one subset. Although height and weight were not used directly as stratification variables, their distributions were balanced across subsets (training: 10.99 ± 3.16 kg, 84.86 ± 13.93 cm; validation: 12.19 ± 3.59 kg, 90.10 ± 14.26 cm; internal test: 12.67 ± 4.37 kg, 90.83 ± 13.97 cm), with one-way ANOVA confirming no significant differences in weight, height, or age across subsets (all $p > 0.05$).

Beyond continuous estimation, predicted height values were converted into height-for-age Z-scores and mapped to the three WHO/Kemenkes categories (normal, stunted, and severely stunted) so that the system could be evaluated as a screening classifier. Classification performance is reported as a confusion matrix with sensitivity, specificity, precision, and F1-score for each category, together with overall accuracy and macro-averaged F1. All principal metrics are accompanied by bias-corrected 95% confidence intervals obtained by bootstrap resampling with 10,000 replications. Agreement between predicted and manually measured values was additionally assessed by Bland–Altman analysis [20], [47], reporting mean bias with its confidence interval, 95% limits of agreement, and regression-based assessment of proportional bias.

In addition to the three error metrics, the strength of linear association between predicted and measured values was assessed using the Pearson correlation coefficient, defined in Eq. (24) [48],

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (24)$$

with r the correlation coefficient, y_i and $\hat{y}_i$ the measured and predicted values, and $\bar{y}$ and $\bar{\hat{y}}$ their respective means across the n samples. The coefficient of determination is defined in Eq. (25) [42], [48],

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (25)$$

with y_i , $\hat{y}_i$ and $\bar{y}$ as defined for Eq. (24). R^2 is the proportion of variance in the measured values explained by the model; a value of 1 indicates perfect prediction, and a value of 0 or below indicates no explanatory power beyond predicting the mean. These statistics are reported alongside the error metrics in Section III.F.

K. Research Ethics

Because this research involves children as subjects, several ethical considerations were carefully addressed throughout the study. Before taking photographs or collecting data from children, written

informed consent was obtained from the parents or guardians, ensuring their understanding and agreement regarding the use of the data. Data confidentiality was strictly maintained; children's identities were not directly linked to the collected data, and all images were stored securely to prevent misuse or unauthorized access. This study received ethical approval from the Medical and Health Research Ethics Committee, Faculty of Medicine, Al-Azhar Islamic University (UNIZAR) Mataram, under Ethical Clearance No. 013/EC-04/FK-06/UNIZAR/I/2026, confirming that the research protocol complies with ethical principles for research involving human subjects.

III. Result

A. Dataset Collection

The dataset consists of 1,593 images from 117 children collected from 8 posyandu in Jenepono Regency. The health centers include Bontoa Village, Pabentengan Village, Jenetallasa Village, Lengke-lengkes Village, Bonia Village, Lianga Village, Kunjungmangge Village, and Bontomanai Village. Each image is accompanied by labels for weight (kg) and height/length (cm) as ground truth results from manual measurements using a digital scale and infantometer/microtois.

B. Preprocessing

All 1,593 images were resized to a fixed dimension of 128×128 pixels before being used as input to the CNN model. In parallel, MediaPipe Pose was applied to each image to extract anatomical landmarks, from which seven predictor features were derived: five inter-landmark distances expressed on a normalized scale of 0 to 1: torso length, upper leg length, lower leg length, total leg length, and total visible height together with two derived ratio features, namely the ratio of torso length to leg length and the ratio of upper leg length to lower leg length. The measured height in centimeters served as the regression target for each image. Because each child contributed 12–15 images captured while the limbs were in natural motion, images from a given child share the same target label and differ only slightly in their feature values. All data partitioning was therefore performed at the child level rather than the image level.

C. Results of the Body Weight Estimation Model Training (CNN)

Table 5 shows that the CNN Model Training has a total of 3,304,769 parameters, all of which are learned during the training process, with no non-trainable parameters. Most of the parameters come from the fully connected layer (Dense 128), which contains over 3 million parameters. This is because this layer connects all 25,088 flattened units to 128 neurons fully.

Table 5. CNN architecture parameters, detailing the layer configuration, output shapes, and parameter counts for each convolutional, pooling, and dense layer.

Layer (type)	Output Shape	Param
conv2d (Conv2D)	(None, 126, 126, 32)	896
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_1 (Conv2D)	(None, 61, 61, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 64)	0
conv2d_2 (Conv2D)	(None, 28, 28, 128)	73,856
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 128)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 128)	3,211,392
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

The basic CNN model was trained using the architecture described in Section II, with the Adam optimizer and Mean Squared Error (MSE) loss function. Training ran for the full 30 epochs; the early-stopping criterion (patience of five epochs on validation loss) was not triggered because validation loss reached a new minimum at least once in every five-epoch window. The training, validation, and test loss curves are shown in Fig. 4, where the dashed line marks the final test loss (MSE = 0.4306). In the graph, the training loss drops sharply in the first few epochs (from around 24 to below 5), indicating that the model quickly learned basic patterns from the data. The validation loss also decreases in a similar pattern, though it shows significant fluctuations around epochs 12–13 (peaking at approximately 4.3) before stabilizing and converging close to the test loss line in the final epochs. These fluctuations are normal for a relatively small dataset and indicate that the model experienced some instability during training but managed to regain stability towards the end of the process.

On the internal test set, the CNN reached an MSE of 0.4306, an RMSE of 0.6562, and an MAE of 0.4923 kg. This internal figure is not an estimate of field accuracy: on the external cohort the MAE was 1.771

kg, which did not improve on an age-only regression (Section III.F). Hyperparameter tuning via grid search tested filter counts in the convolutional layers, dense units, and dropout rate. The best candidate (filters1 = 32, filters2 = 128, dense_units = 256, dropout = 0.3) reduced validation error only marginally, by about 0.019 in MAE (0.4737 kg), 0.001 in MSE, and 0.0008 in RMSE, and when retrained, its test loss (0.5138) was higher than that of the base model (0.4306). Because the tuned candidate did not improve test performance and would have doubled the size of the dense layer, the base configuration with a 128-unit dense layer, described in Section II and Table 3, was retained as the final model and used for all external evaluation. Scatter plots of actual against predicted weight for the baseline and tuned models on the validation data were visually indistinguishable, consistent with the negligible difference in error metrics, and are therefore not reproduced here.

D. Results of the Body Height Estimation Model Training (MediaPipe Pose + Regression)

The Random Forest Regressor model was trained using features extracted from MediaPipe Pose (torso_length, upper_leg_length, lower_leg_length, leg_length, total_visible_height, and two derived ratios) as predictor variables, with actual height (height_cm) as the target. To improve performance, hyperparameter tuning was also performed on the Random Forest model. On the internal validation set, the baseline Random Forest achieved an MAE of 5.231 cm and the tuned model an MAE of 5.224 cm, so tuning improved performance only marginally (an MAE reduction of 0.007 cm, with reductions of about 0.50 in MSE and 0.03 in RMSE). This pattern is similar to what was found in the CNN weight estimation model indicating that the basic model configuration is already relatively optimal for the pose feature characteristics used, so the scope for improvement through hyperparameter tuning is limited. One of the advantages of Random Forest is its ability to show the contribution level (importance) of each feature to the prediction results. The leg_length feature (total length

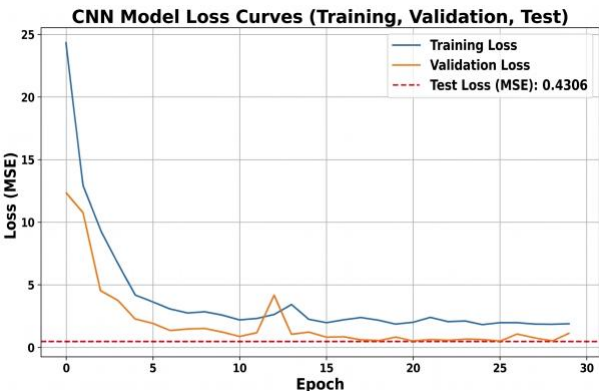


Fig. 4. Training, validation, and test loss curves of the CNN model over 30 epochs, showing convergence with a final test loss (MSE) of 0.4306.

of the leg, from the hip to the ankle) provides the most dominant contribution to height prediction, with an importance score of 0.4339 much higher than other features. This aligns with anthropometric principles that leg length has a strong correlation with a person's total height. The `total_visible_height` feature (height visualized from the top of the head to the feet) ranks second, which is reasonable because this feature directly represents a rough proxy of the height itself. Interestingly, the `upper_lower_leg_ratio` feature also contributes quite significantly (0.1108), higher than `lower_leg_length` itself, indicating that the proportion between the thigh and calf also helps the model in correcting height estimates, especially to account for variations in a child's posture during image capture.

The two models produced visually indistinguishable prediction patterns, consistent with the MAE difference of approximately 0.007 cm between them. In both cases, predictions clustered closely around the ideal line in the 85–100 cm range, where the dataset is densest, and deviated more below 70 cm and above 100 cm, where fewer samples were available; these plots are therefore not reproduced here.

Prediction API whenever a new image is uploaded for processing.

To assess the feasibility of the system used in the field, a processing time test was conducted from the moment the image is uploaded until the nutritional status classification results are displayed to the user. Both the models and the FastAPI service are hosted on Hugging Face Spaces with 2 vCPU and 16 GB RAM and no GPU acceleration; no inference is performed on the client device, so the pipeline consists of network transfer to the remote endpoint followed by remote computation. Response time was measured across 48 trials over a cellular connection with nominal throughput of 40 to 70 Mbps, using photographs of 2 to 4 MB. Table 6 reports the distribution of each component. Mean end-to-end response time was 1.143 ± 0.499 s (median 1.295 s, IQR 0.540–1.600 s), of which image upload took 0.769 ± 0.496 s, model inference 0.239 ± 0.017 s, and status computation 0.134 ± 0.015 s.

Latency is dominated by network transfer rather than computation. Image upload accounts for 67.3% of

Table 6. Distribution of response-time components across 48 deployment trials on Hugging Face Spaces (2 vCPU, 16 GB RAM, CPU-only), measured over a cellular connection at 40–70 Mbps with image payloads of 2–4 MB.

Component	Mean ± SD (s)	Median (IQR) (s)	Range (s)	CV (%)
Image upload	0.769 ± 0.496	0.920 (0.140–1.232)	0.080–1.380	64.4
Model inference	0.239 ± 0.017	0.240 (0.230–0.250)	0.200–0.280	7.3
Status computation	0.134 ± 0.015	0.130 (0.128–0.140)	0.110–0.180	11.4
End-to-end total	1.143 ± 0.499	1.295 (0.540–1.600)	0.450–1.810	43.7

E. System Integration

CNN model for weight estimation and a MediaPipe Pose-based Random Forest model for height estimation are integrated into a system called AI-StuntDetect, with system responsibilities separated into two main layers. First, a Python-based Prediction API that wraps these machine learning models and exposes them as REST API endpoints to receive images as input and return the estimated weight (kg) and height (cm) results in JSON format. Second, a web application based on Laravel and PostgreSQL that serves as the main interface for Posyandu cadres, handling multi-role user management, storing child data and measurement history, and calling the

mean end-to-end time and 98.8% of its variance; end-to-end time correlates almost perfectly with upload time ($r = 0.999$, $p < 0.001$) while its correlation with inference time is not significant ($r = 0.258$, $p = 0.077$). The two computational components are highly stable, neither exceeding 0.28 s in any trial.

F. Manual Validation

Testing was conducted on 48 children recruited across three independent posyandu in Jeneponto (Mangambang, Belokallong, and Karisa), none of whom contributed data to any stage of model development, to assess the accuracy of the image-based height and weight estimation models, including

Table 7. Model performance against baseline predictors on the external cohort (n = 48). The proposed model outperforms every baseline for height estimation, but does not improve on the age-only regression for weight estimation.

Predictor	Height MAE (cm)	Height RMSE (cm)	Weight MAE (kg)	Weight RMSE (kg)
Test-cohort mean	11.165	13.170	2.715	3.235
Training-set mean	11.118	13.230	2.719	3.249
Age-only regression	6.005	7.896	1.675	2.256
Age + sex regression	5.923	7.941	1.711	2.394
AI-StuntDetect	4.092	4.938	1.771	2.222

their performance against the baseline predictors in Table 7. Each sample was tested by comparing manual measurements with the model's predictions. The differences were calculated as the difference between actual and predicted values, and their agreement is summarized in the Bland–Altman plots of Fig. 5; positive values indicate the model tends to underestimate, while negative values indicate an overestimate.

Model accuracy evaluation was conducted using three error metrics: MAE, MSE, and RMSE. MAE indicates the average magnitude of errors regardless of the direction of deviation, MSE places greater weight on extreme errors because the differences are squared, while RMSE converts the MSE back to the original units, making it easier to interpret. For height estimation the model achieved an MAE of 4.092 cm (95% CI 3.321–4.888), an MSE of 24.386 cm², and an RMSE of 4.938 cm (95% CI 4.022–5.841). For weight estimation the model achieved an MAE of 1.771 kg (95% CI 1.406–2.156), an MSE of 4.936 kg², and an

RMSE of 2.222 kg (95% CI 1.786–2.640).

Table 8. Confusion matrix for three-category stunting classification on the external validation cohort (n = 48), comparing categories derived from model-predicted height against manual measurement using the WHO/Kemenkes height-for-age Z-score reference standard.

Actual \ Predicted	Normal	Stunted	Severely Stunted
Normal	29	7	1
Stunted	3	1	2
Severely Stunted	1	0	4

Predicted height values were converted to height-for-age Z-scores and mapped to the three

Table 9. Per-class classification performance on the external validation cohort (n = 48). Overall accuracy was 0.708 (95% CI 0.583–0.833) with a macro-averaged F1 of 0.546 (95% CI 0.373–0.693), computed one-versus-rest for each category.

Class	n	Sensitivity	Specificity	Precision	F1-score
Normal	37	0.784	0.636	0.879	0.829
Stunted	6	0.167	0.833	0.125	0.143
Severely Stunted	5	0.800	0.930	0.571	0.667

Descriptive statistics for the external cohort were as follows: measured height ranged from 56.8 to 107.0 cm (mean 83.75 cm; SD 13.32 cm) and measured weight from 4.7 to 18.2 kg (mean 10.73 kg; SD 3.27 kg). Predicted values spanned 60.00 to 101.08 cm for height (SD 11.67 cm) and 5.20 to 18.70 kg for weight (SD 2.75 kg), representing 87.6% and 84.0% of the observed dispersion respectively. Agreement between predicted and measured values was strong for height ($r = 0.929$; $R^2 = 0.863$, 95% CI 0.787–0.917) and moderate for weight ($r = 0.768$; $R^2 = 0.590$, 95% CI 0.364–0.761). Both models therefore reproduce a substantial proportion of the variation present in the cohort, and neither exhibits the range compression characteristic of a model defaulting to the sample mean.

Table 7 reports performance against each baseline. Against the mean-value baseline the model reduced height MAE by 63.3% and weight MAE by 34.8%, so both models clearly outperform a constant predictor. Against the training-derived demographic baselines the picture is asymmetric. For height the model reduced MAE by a further 31.9% relative to the strongest baseline, from 5.923 cm to 4.092 cm. For weight the model did not improve on the age-only

WHO/Kemenkes categories. Table 8 presents the confusion matrix and Table 9 the corresponding per-class metrics. Of the 48 children, 34 were classified correctly: 29 of 37 normal, 1 of 6 stunted and 4 of 5 severely stunted. Overall accuracy was 0.708 (95% CI 0.583–0.833) with a macro-averaged F1 of 0.546 (95% CI 0.373–0.693). Performance was strongest at the extremes of the nutritional spectrum: normal status was identified with a sensitivity of 0.784, a precision of 0.879 and an F1 of 0.829, while severely stunted cases were detected with a sensitivity of 0.800, a specificity of 0.930 and an F1 of 0.667. Treating stunting of any severity as a single positive class, 7 of the 11 affected children (63.6%) were flagged as requiring further assessment and 4 were classified as normal (3 stunted and 1 severely stunted), while 8 of the 37 normal children were flagged incorrectly, 7 as stunted and 1 as severely stunted. Performance was markedly lower for the intermediate stunted category, with a sensitivity of 0.167, a precision of 0.125 and an F1 of 0.143, and the misclassifications there were predominantly adjacent-category errors: of the five stunted children not correctly labelled, three were assigned to the normal category and two to the severely stunted category. This is consistent with the magnitude of the underlying height error, since the stunted category occupies a

narrow Z-score band whose width is comparable to the residual measurement error of the estimation pipeline.

three largest all occurred in children above 100 cm in true stature, the region where the proportional bias

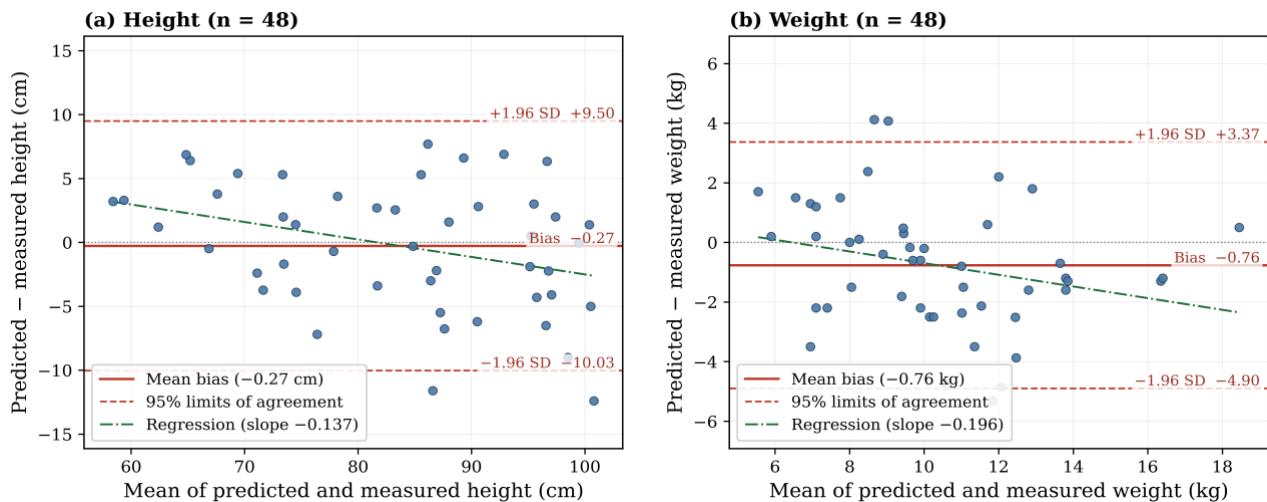


Fig. 5. Bland–Altman agreement plots for height (a) and weight (b) on the external validation cohort (n = 48), showing mean bias (solid red), 95% limits of agreement (dashed red), and the regression of differences on their means (dash-dotted green).

Agreement beyond correlation was assessed by Bland–Altman analysis, shown in Fig. 5. For height the mean bias was -0.27 cm (95% CI -1.68 to $+1.14$), not significantly different from zero ($p = 0.713$), with 95% limits of agreement of -10.03 to $+9.50$ cm and a small but significant proportional bias (slope -0.137 , $p = 0.019$). For weight the mean bias was -0.76 kg (95% CI -1.36 to -0.17), significantly different from zero ($p = 0.016$), with limits of agreement of -4.90 to $+3.37$ kg and no significant proportional bias (slope -0.196 , $p = 0.071$).

Error was stratified by age group and by conventional measurement mode. Absolute error did not differ significantly between the recumbent and standing regimes (3.681 cm below 24 months versus 4.362 cm at and above, $p = 0.371$), but the direction did: the model over-estimated stature by 1.87 cm on average below 24 months and under-estimated it by 1.66 cm at and above, a significant difference in signed bias ($p = 0.010$). Across the full range, signed height error correlated negatively with age ($r = -0.386$, $p = 0.007$) while absolute error did not ($r = 0.070$, $p = 0.638$), indicating an age-dependent systematic bias rather than an age-dependent loss of precision. Weight bias followed the same direction ($+0.49$ kg below 24 months, -1.59 kg at and above).

Errors were not uniformly distributed: 31 of 48 predictions fell within 5 cm of the manual measurement while 2 exceeded 10 cm. The median absolute height error was 3.50 cm (IQR 2.00 – 6.24), below the mean of 4.09 cm, indicating that the mean is inflated by a small number of large failures. Four of the five largest errors occurred in children of 24 months or older, and the

identified above is strongest; all three were under-estimations. The failures are therefore directionally consistent rather than random, pointing to a systematic limitation of the pose-to-stature mapping at the upper end of the measured range rather than to sporadic landmark detection failure. Children whose nutritional category was classified correctly had a mean absolute height error of 3.31 cm, against 5.99 cm among those misclassified ($p = 0.012$).

IV. Discussion

This study found that the proposed CNN and MediaPipe Pose system estimates height with useful accuracy (MAE 4.092 cm; RMSE 4.938 cm) and estimates weight with moderate accuracy (MAE 1.771 kg; RMSE 2.222 kg), but that the two results carry different implications once the choice of baseline is made more demanding. Both models comfortably outperform a constant mean-value predictor, by 63.3% and 34.8% in MAE respectively. Only the height model, however, outperforms a demographic predictor estimated from the training data: an age-only linear regression achieves 1.675 kg on the external cohort, marginally better than the image-based weight model, whereas the strongest demographic baseline for height reaches only 5.923 cm against the model's 4.092 cm. These findings indicate that a single 2D image combined with pose-landmark-based feature extraction is sufficient to estimate linear anthropometric measures such as height, but is inadequate for mass-dependent measures such as weight, which appear to require volumetric or depth information that a monocular 2D image cannot provide.

This distinction was not visible in our original evaluation, which used a cohort of 13 children. On that sample the weight model showed a correlation of 0.056 with measured values and a predicted range spanning only 8.31 to 11.19 kg, which we interpreted as evidence that it had learned nothing beyond the sample mean. The expanded cohort of 48 children does not support that interpretation: correlation rises to 0.768, R^2 to 0.590, and the predicted range widens to 5.20 to 18.70 kg, 84.0% of the dispersion actually present. The weight model does learn genuine subject-specific variation. What it does not do is improve on what age alone predicts, and this is a more demanding and more informative test than the mean-value comparison we originally applied. We report the earlier interpretation and its revision explicitly, because the episode illustrates the instability of conclusions drawn from small external samples, which is itself a finding of practical value to the field.

The CNN attained an MAE of 0.4923 kg on the internal partition against 1.771 kg externally. Because that partition was split at the child level, the degradation cannot be attributed to identity leakage. Formal comparison of the two cohorts shows they are statistically indistinguishable in height ($p = 0.252$), weight ($p = 0.246$), and sex ($p = 0.893$), but differ in age composition, both in central tendency (33.5 versus 27.4 months, $p = 0.032$) and in distributional shape ($KS = 0.281$, $p = 0.007$). Since model error is itself age-dependent, as the age-stratified analysis shows, a cohort with different age composition will exhibit different aggregate error. Of the acquisition variables, camera geometry and mat placement were held constant across both phases and cannot contribute; ambient lighting and non-standardized clothing did vary, and clothing in particular affects the silhouette from which body mass is inferred far more than it affects the skeletal landmarks used for stature. The internal figures therefore describe performance on children from the same sites and collection period as the training data and should not be read as estimates of field accuracy; the external results are the relevant measure for deployment. The height model shows the opposite pattern, with an internal MAE of 5.224 cm on the 19 validation children against 4.092 cm externally. Given the small internal partitions, we attribute this to sampling variability rather than to better generalization, and neither internal figure is directly comparable with the external results.

Table 10 places AI-StuntDetect alongside six related studies, compared by population, sample size, acquisition modality, validation strategy, and reported error. No prior study matches ours on all these dimensions, so the comparison provides context rather than a ranking. The studies fall into three groups: CT-based estimation, RGB estimation on subjects with

restricted posture, and keypoint-based regression. Each group differs from ours along a different dimension.

The CT-based studies [23], [24] differ from ours mainly in population and modality. Ichikawa et al. [23] estimated weight from CT scout images of 1,831 (chest) and 519 (abdominal) adults aged 24–92 years, reporting MAEs of 2.75 and 4.77 kg under five-fold cross-validation with a 20% hold-out test. Demircioğlu et al. [24] estimated height and weight from CT localizers of pediatric and young adult patients aged 1 month to 21 years and, on a temporally separated test set, reported 5.58 ± 4.26 cm and 4.25 ± 4.28 kg. Their subjects span a body-size range several times wider than that of children under five, which inflates absolute error. In relative terms, the height errors are of the same order: [24] reported a mean relative error of 4.12%, and our MAE equals 4.9% of the external cohort's mean height of 83.8 cm. CT also captures volumetric information that a single RGB photograph cannot, which matters most for weight.

The RGB studies [27], [29], [30] share our modality and report lower height errors of 1.1, 1.52, and 2.28 cm, but under conditions that remove the main sources of error in our setting. Shu et al. [27] photographed 200 newborns, whose movement is restricted and whose size range is narrow, and evaluated by cross-validation within the same collection. Studies [29], [30] measured physical doll models with fixed postures and validated them in controlled laboratory testing. Our external cohort spans 2 to 67 months, an interval over which body proportions change and postural cooperation varies from child to child, and all images were taken of real children at posyandu sites.

The study closest to ours in pipeline design is [28], which converted YOLO-detected body keypoints into geometric features and regressed height with AdaBoost, reporting an MAE of 6.2 cm on 4,189 images. Our pipeline follows the same keypoint-to-feature-to-ensemble structure, using MediaPipe Pose and a Random Forest, and reaches 4.092 cm. We do not attribute this difference to the choice of detector or regressor, because the two studies differ in population and validation design, and our fixed acquisition geometry removes the principal source of scale ambiguity in monocular height estimation.

Validation design is the dimension on which our study differs most from the others. The comparison studies report internal cross-validation [23], [27], a temporally separated test set from the same source [24], or laboratory testing [29], [30]. Our figures come from 48 children at three posyandu that contributed no training data. External evaluation of this kind exposes site-specific effects that internal splits cannot detect, so the errors in Table 10 were obtained under different levels of stringency. We report RMSE (4.938 cm for

height and 2.222 kg for weight) alongside MAE, because RMSE penalizes the large individual errors that matter in screening, to support future comparisons.

achieves 1.675 kg on the same cohort, so a low absolute error in this population is attainable without any image information. Because none of the

Table 10. Comparison of AI-StuntDetect with six related anthropometric estimation studies by population, sample size, acquisition modality and method, validation strategy, and reported error. Errors are not directly comparable because the studies differ in age range (newborns to adults), subject type (children, patients, physical models), modality (single RGB photograph versus volumetric CT), and validation design (internal cross-validation to external multi-site testing).

Study	Population	n	Imaging modality	Validation strategy	Target	Reported error
This study	Indonesian children; development 0–60 mo, external 2–67 mo	117 dev (1,593 images) / 48 ext	Single RGB photograph, fixed geometry; CNN + MediaPipe Pose + RF	External, 3 independent posyandu	Height	MAE 4.09 cm (95% CI 3.32–4.89); RMSE 4.94 cm
This study	Indonesian children; development 0–60 mo, external 2–67 mo	117 dev (1,593 images) / 48 ext	Single RGB photograph, fixed geometry; CNN + MediaPipe Pose + RF	External, 3 independent posyandu	Weight	MAE 1.77 kg (95% CI 1.41–2.16); RMSE 2.22 kg
[24]	Pediatric and young adult patients, 1 mo–21 y	1,009–1,111 train; 116–127 val; 203–225 test	CT localizer; EfficientNetV2-S	Temporal hold-out (train 2013–2019, test 2021–2023)	Height; Weight	MAE 5.58 ± 4.26 cm; MAE 4.25 ± 4.28 kg
[23]	Adult patients, 24–92 y	1,831 chest; 519 abdominal	CT scout image; CNN	5-fold CV with 20% hold-out test	Weight	MAE 2.75 kg (chest); 4.77 kg (abdominal)
[28]	Adults	4,189 images	RGB image; YOLO keypoints + AdaBoost	80% Training and 20% Testing	Height	MAE 6.2 cm
[27]	Newborns	200	Single RGB photograph; CNN (MobileNet, VGG16, GoogleNet, AlexNet)	5-fold cross-validation	Height; Weight	MAE < 1.1 cm; MAE < 0.28 kg
[29]	Dolls (physical models)	12 dolls	RGB photograph; MediaPipe Pose	Multi-stage Validation	Height	MAE 1.52 cm
[30]	Dolls (physical models)	2 (infant mannequins, 50 and 60 cm; 44 measurements)	RGB photograph; pose estimation (model pre-trained on MPII)	Controlled Laboratory Simulation Testing	Height	MAE 2.28 cm

For weight, the comparison requires particular care. Our MAE of 1.771 kg is lower than the 2.75–4.77 kg of [23] and the 4.25 kg of [24], and higher than the 0.28 kg reported for newborns in [27]. In each case, the difference follows the weight range of the population more closely than any property of the method. Our baseline analysis makes this point concretely: an age-only regression fitted to the training data alone

comparison studies reports baseline-adjusted performance, their figures and ours cannot be placed on a common scale, and we claim no superiority over them for weight.

The classification results need separate interpretation, because overall accuracy (0.708) and macro-F1 (0.546) conceal a class-specific weakness

that matters most for early screening. Sensitivity for the stunted category was only 0.167 (F1 0.143): of the six stunted children, one was classified correctly, three were classified as normal, and two as severely stunted. Across both affected categories, 4 of 11 children were classified as normal. In practice, a child whose stunting is missed would not be referred for re-measurement or nutritional counselling at that session, and the intermediate stunted category is where early intervention is most useful, before growth faltering becomes severe. The weakness follows from the size of the height error relative to the width of the category: the stunted band spans one Z-score unit, equivalent to about 2.5 to 4.5 cm of height between 1 and 5 years of age, which is comparable to the model's MAE of 4.092 cm, so a moderate error is enough to move a child across a category boundary. Two consequences follow. A normal result from the system cannot be used to rule out stunting, so routine manual measurement must continue for all children; and the system is better suited to flagging children for confirmation than to assigning nutritional status. Referring every child whose predicted Z-score lies within a margin of the -2 threshold for manual re-measurement would trade additional re-measurements for fewer missed cases; the appropriate margin should be determined in the larger validation cohort.

Six limitations define the scope of these results, and each sets a specific requirement for future validation. First, all three external sites lie within one regency, and only 11 children were stunted or severely stunted, so the classification metrics carry wide confidence intervals and are preliminary; validation therefore requires a multi-regency cohort large enough to estimate sensitivity for the stunted category with adequate precision. Second, camera device, camera height, overhead-arm geometry, mat dimensions, and subject posture were fixed in both phases, so the reported accuracy is an upper bound for unconstrained capture; the system must be tested across devices, camera-to-subject distances, and viewing angles. Third, reference measurement switched from recumbent to standing at 12 months, following posyandu practice, whereas the WHO standards assume recumbent length until 24 months (an expected difference of about 0.7 cm), and all images were captured supine; this mismatch plausibly contributes to the age-dependent bias, so future reference data should use recumbent length below 24 months, matching the imaging posture. Fourth, each child was measured once, so the reported error combines model error with reference error; repeated measurements by at least two trained cadres are required to estimate inter- and intra-rater reliability and separate the two sources.

Fifth, results derive from a single training run, so run-to-run variability is not quantified; repeated training across multiple random seeds, reported as mean $\pm$ standard deviation, is required, and replacing the flatten-to-dense transition with global average pooling (about 110,000 parameters) will be evaluated in the same experiment. Sixth, response time was measured on one cellular connection, and inference requires network access; timing under the weaker connectivity of rural posyandu must be characterized, and an offline, on-device mode evaluated.

Practically, these results support a hybrid workflow in which height is estimated from the image while weight continues to be measured with conventional scales. The justification for this split is now sharper than a simple accuracy comparison: the height model provides information demographic data alone cannot supply, whereas the weight model does not, so automating weight estimation would add system complexity and a connectivity dependency without a corresponding gain.

The deployment analysis adds a further practical note. Since 67.3% of response time is network transfer and the computational components are stable to within hundredths of a second, client-side image compression rather than model optimization is the effective route to improved responsiveness in the field.

A second methodological implication follows from the revision of our own conclusion. Our original evaluation, conducted on 13 children, indicated that the weight model had learned nothing beyond the sample mean; on 48 children the same model explains 59% of the variance. Conclusions about model behaviour drawn from external cohorts of this size are unstable in both directions, and the practice of reporting confidence intervals alongside point estimates, which we have adopted throughout the revised analysis, would make such instability visible at review. Equally, the choice of baseline determines what a comparison can reveal: a mean-value baseline was insufficient to identify that our weight model, while genuinely predictive, adds nothing to a trivial age regression.

V. Conclusion

This study developed and externally evaluated AI-StuntDetect, a web-based system that estimates height and weight from a single photograph to support stunting screening at posyandu. On 48 children from three independent posyandu, height estimation achieved an MAE of 4.092 cm ($R^2 = 0.863$), 31.9% lower than the strongest demographic baseline, with negligible group-level bias (-0.27 cm). Weight estimation achieved an MAE of 1.771 kg but did not improve on an age-only regression (1.675 kg). Stunting classification reached an accuracy of 0.708 and a macro-F1 of 0.546, with a

sensitivity of only 0.167 for the stunted category, and the mean response time was 1.143 s.

These results support a hybrid workflow in which height is estimated from the image while weight continues to be measured with a scale. Because the 95% limits of agreement for height (-10.03 to $+9.50$ cm) are far wider than manual measurement error, and 4 of the 11 affected children were classified as normal, the system in its current form can assist but not replace routine anthropometry, and every screening result requires confirmation by a trained cadre. The major limitation is the size and geographic range of the external cohort, which came from a single regency and included only 11 stunted or severely stunted children.

Future work will (1) validate the system in a larger multi-regency cohort sized to estimate stunted-class sensitivity with adequate precision, (2) test it under unconstrained capture across devices, distances, and viewing angles, (3) reduce the age-dependent bias through age-conditioned calibration against recumbent reference measurements below 24 months, and (4) report multi-seed training results and evaluate an offline mode for posyandu with weak connectivity.

Acknowledgment

The authors thank the Ministry of Science and Technology, Indonesia, and the Directorate of Research and Community Service for supporting this work through the Doctoral Dissertation Research Grant Program.

Funding

This research was funded by the Ministry of Science and Technology, Indonesia, through the Directorate of Research and Community Service under the Doctoral Dissertation Research Grant Program for the 2026 Fiscal Year, contract number 2803/DST/UN36.12/TU/2026.

Data Availability

The datasets generated and analyzed during the study are not publicly available because they contain images of minors and are subject to privacy and ethical restrictions. De-identified data may be made available from the corresponding author upon reasonable request and appropriate ethical approval.

Author Contributions

Asmaul Husna RS contributed to conceptualization, methodology, data curation, software development, formal analysis, validation, investigation, and preparation of the original draft. Purnamawati contributed to supervision, conceptualization, methodology, review, and editing. Hendra Jaya and

Anas Arfandi contributed to supervision, review, and editing.

Declarations

Ethics Approval

This study received ethical approval from Medical and Health Research Ethics Committee, Faculty of Medicine, Al-Azhar Islamic University (UNIZAR) Mataram, Indonesia, approval number 013/EC-04/FK-06/UNIZAR/II/2026.

Informed Consent to Participate

Written informed consent was obtained from the parents or legal guardians of all participating children before enrollment and data collection.

Consent for Publication

No identifiable participant information is published in this article. Consent covering participation and the use of the collected images for research purposes was obtained from parents or guardians.

Competing Interests

The authors declare that they have no competing financial interests or personal relationships that could have influenced the work reported in this manuscript.

Authors' Approval

All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work.

Availability of Code and Models

The source code, trained models, and deployment scripts developed in this study are not publicly available because they form part of an ongoing doctoral research project and a system under further development. They are available from the corresponding author upon reasonable request.

References

- [1] S. Rahardianto *et al.*, "Spatio-temporal modeling to identify factors associated with stunting in Indonesia using a modified generalized lasso," *Spat. Spatiotemporal Epidemiol.*, vol. 51, Art. no. 100694, 2024, doi: [10.1016/j.sste.2024.100694](https://doi.org/10.1016/j.sste.2024.100694).
- [2] P. K. Arya, K. Sur, T. Kundu, S. Dhote, and S. K. Singh, "Unveiling predictive factors for household-level stunting in India: A machine learning approach using NFHS-5 and satellite-driven data," *Nutrition*, vol. 132, Art. no. 112674, 2025, doi: [10.1016/j.nut.2024.112674](https://doi.org/10.1016/j.nut.2024.112674).
- [3] A. Ndagijimana, G. Nduwayezu, T. Lind, and A. Mansourian, "Machine learning techniques to model child low height-for-age in the Northern

- Province of Rwanda: The role of climatological and environmental factors and their interactions," *Clin. Epidemiol. Glob. Health*, vol. 37, Art. no. 102284, 2026, doi: [10.1016/j.cegh.2025.102284](https://doi.org/10.1016/j.cegh.2025.102284).
- [4] V. G. Biradar, H. H. P. K. Pareek, S. T. K. Aishwarya, and K. Aneeth, "Performance analysis of deep learning models for detection of malnutrition in children," in *Proc. 2024 Int. Conf. Data Sci. Netw. Secur. (ICDSNS)*, Jul. 2024, pp. 1–8, doi: [10.1109/ICDSNS62112.2024.10691135](https://doi.org/10.1109/ICDSNS62112.2024.10691135).
 - [5] S. Jain, T. Khanam, A. J. Abedi, and A. A. Khan, "Efficient machine learning for malnutrition prediction among under-five children in India," in *Proc. 2022 IEEE Delhi Sect. Conf. (DELCON)*, 2022, pp. 1–10, doi: [10.1109/DELCON54057.2022.9753080](https://doi.org/10.1109/DELCON54057.2022.9753080).
 - [6] N. Akseer *et al.*, "Economic costs of childhood stunting to the private sector in low- and middle-income countries," *EClinicalMedicine*, vol. 45, Art. no. 101320, 2022, doi: [10.1016/j.eclinm.2022.101320](https://doi.org/10.1016/j.eclinm.2022.101320).
 - [7] I. Mostafa *et al.*, "A microbiota-directed complementary food intervention in 12–18-month-old Bangladeshi children improves linear growth," *eBioMedicine*, vol. 104, Art. no. 105166, 2024, doi: [10.1016/j.ebiom.2024.105166](https://doi.org/10.1016/j.ebiom.2024.105166).
 - [8] Md. M. Islam *et al.*, "Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh," *Int. J. Cogn. Comput. Eng.*, vol. 3, pp. 46–57, 2022, doi: [10.1016/j.ijcce.2022.02.002](https://doi.org/10.1016/j.ijcce.2022.02.002).
 - [9] R. E. Kofinti, I. Koomson, J. A. Paintsil, and E. K. Ameyaw, "Reducing children's malnutrition by increasing mothers' health insurance coverage: A focus on stunting and underweight across 32 sub-Saharan African countries," *Econ. Model.*, vol. 117, Art. no. 106049, 2022, doi: [10.1016/j.econmod.2022.106049](https://doi.org/10.1016/j.econmod.2022.106049).
 - [10] E. Gayawan and O. A. Egbon, "Spatio-temporal mapping of stunting and wasting in Nigerian children: A bivariate mixture modeling," *Spat. Stat.*, vol. 58, Art. no. 100785, 2023, doi: [10.1016/j.spasta.2023.100785](https://doi.org/10.1016/j.spasta.2023.100785).
 - [11] S. V. Subramanian, O. Karlsson, and R. Kim, "Revisiting the stunting metric for monitoring and evaluating nutrition policies," *Lancet Glob. Health*, vol. 10, no. 2, pp. e179–e180, 2022, doi: [10.1016/S2214-109X\(21\)00504-0](https://doi.org/10.1016/S2214-109X(21)00504-0).
 - [12] L. Bevis, K. Kim, and D. Guerená, "Soil zinc deficiency and child stunting: Evidence from Nepal," *J. Health Econ.*, vol. 87, Art. no. 102691, 2023, doi: [10.1016/j.jhealeco.2022.102691](https://doi.org/10.1016/j.jhealeco.2022.102691).
 - [13] B. Haile and D. Headey, "Growth in milk consumption and reductions in child stunting: Historical evidence from cross-country panel data," *Food Policy*, vol. 118, Art. no. 102485, 2023, doi: [10.1016/j.foodpol.2023.102485](https://doi.org/10.1016/j.foodpol.2023.102485).
 - [14] UNICEF, WHO, and World Bank, "Levels and trends in child malnutrition: UNICEF/WHO/World Bank Group joint child malnutrition estimates: Key findings of the 2023 edition," World Health Organization, Geneva, Switzerland, 2023. ISBN: 978-92-4-007379-1.
 - [15] R. D. Logarajan, N. M. Nor, S. Ibrahim, and R. Said, "Social determinants of stunting in Malay children aged <5 years in Malaysia," *Nutrition*, vol. 111, Art. no. 112030, 2023, doi: [10.1016/j.nut.2023.112030](https://doi.org/10.1016/j.nut.2023.112030).
 - [16] M. B. Arndt *et al.*, "Global, regional, and national progress towards the 2030 global nutrition targets and forecasts to 2050: A systematic analysis for the Global Burden of Disease Study 2021," *Lancet*, vol. 404, no. 10471, pp. 2543–2583, 2024, doi: [10.1016/S0140-6736\(24\)01821-X](https://doi.org/10.1016/S0140-6736(24)01821-X).
 - [17] H. A. Santoso, N. S. Dewi, S.-C. Haw, A. D. Pambudi, and S. A. Wulandari, "Enhancing nutritional status prediction through attention-based deep learning and explainable AI," *Intell.-Based Med.*, vol. 11, Art. no. 100255, 2025, doi: [10.1016/j.ibmed.2025.100255](https://doi.org/10.1016/j.ibmed.2025.100255).
 - [18] N. Khalid, R. Nawaz, Z. Zhou, and S. Gong, "Child malnutrition in Pakistan: Pathways linking maternal intimate partner violence, gender inequality, healthcare utilization, and age at marriage," *Child. Youth Serv. Rev.*, vol. 186, Art. no. 109039, 2026, doi: [10.1016/j.childyouth.2026.109039](https://doi.org/10.1016/j.childyouth.2026.109039).
 - [19] K. Simamora, B. W. Otok, and S. Setiawan, "Multilevel confirmatory factor analysis for modelling the determinants of malnutrition among under-five children in Indonesia," in *Proc. 2025 Int. Conf. Artif. Intell., Blockchain, Cloud Comput., Data Anal. (ICoABCD)*, 2025, pp. 1–5, doi: [10.1109/ICoABCD67551.2025.11470742](https://doi.org/10.1109/ICoABCD67551.2025.11470742).
 - [20] R. Mukherjee *et al.*, "Accuracy of a smartphone-based 3D imaging tool for height measurement and stunting detection among children 24–59 months of age in Nepal: A validation study," *Lancet Reg. Health Southeast Asia*, vol. 48, Art. no. 100774, 2026, doi: [10.1016/j.lanseas.2026.100774](https://doi.org/10.1016/j.lanseas.2026.100774).
 - [21] H. Anastasia *et al.*, "Determinants of stunting in children under five years old in South Sulawesi and West Sulawesi Province: 2013 and 2018 Indonesian Basic Health Survey," *PLoS One*, vol. 18, no. 5, Art. no. e0281962, 2023, doi: [10.1371/journal.pone.0281962](https://doi.org/10.1371/journal.pone.0281962).
 - [22] Health Development Policy Agency, Ministry of Health of the Republic of Indonesia, *Indonesian Health Survey 2023 in Numbers*. Jakarta, Indonesia: Ministry of Health of the Republic of Indonesia, 2024.

- [23] S. Ichikawa, M. Hamada, and H. Sugimori, "A deep-learning method using computed tomography scout images for estimating patient body weight," *Sci. Rep.*, vol. 11, no. 1, Art. no. 15627, 2021, doi: [10.1038/s41598-021-95170-9](https://doi.org/10.1038/s41598-021-95170-9).
- [24] A. Demircioğlu, A. S. Quinsten, L. Umutlu, M. Forsting, K. Nassenstein, and D. Bos, "Determining body height and weight from thoracic and abdominal CT localizers in pediatric and young adult patients using deep learning," *Sci. Rep.*, vol. 13, no. 1, Art. no. 19010, 2023, doi: [10.1038/s41598-023-46080-5](https://doi.org/10.1038/s41598-023-46080-5).
- [25] L. D. B. Lima *et al.*, "A scale-equivariant CNN-based method for estimating human weight and height from multi-view clinic silhouette images," *Expert Syst. Appl.*, vol. 256, Art. no. 124879, 2024, doi: [10.1016/j.eswa.2024.124879](https://doi.org/10.1016/j.eswa.2024.124879).
- [26] V. G. Sonar *et al.*, "3D camera-based body weight estimation using artificial intelligence in emergency care settings," *Emerg. Care Med.*, vol. 2, no. 4, Art. no. 55, 2025, doi: [10.3390/ecm2040055](https://doi.org/10.3390/ecm2040055).
- [27] H. Shu, L. Ren, L. Pan, D. Huang, H. Lu, and W. Wang, "Single image based infant body height and weight estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2023, pp. 6052–6059, doi: [10.1109/CVPRW59228.2023.00644](https://doi.org/10.1109/CVPRW59228.2023.00644).
- [28] M. Sakina, I. Muhammad, and S. S. Abdullahi, "A multi-factor approach for height estimation of an individual using 2D image," *Procedia Comput. Sci.*, vol. 231, pp. 765–770, 2024, doi: [10.1016/j.procs.2023.12.140](https://doi.org/10.1016/j.procs.2023.12.140).
- [29] Risfendra, Aripriharta, Suherman, G. F. Ananda, D. S. Putra, and Fahmi, "Contactless infant height measurement for enhanced early detection of stunting using computer vision techniques," *IEEE Access*, vol. 13, pp. 42364–42376, 2025, doi: [10.1109/ACCESS.2025.3548159](https://doi.org/10.1109/ACCESS.2025.3548159).
- [30] L. Qiao *et al.*, "Contactless height measurement for infants and toddlers based on computer vision," in *Proc. 2024 4th Int. Conf. Consum. Electron. Comput. Eng. (ICCECE)*, 2024, pp. 306–310, doi: [10.1109/ICCECE61317.2024.10504244](https://doi.org/10.1109/ICCECE61317.2024.10504244).
- [31] P. Nguyen Trung *et al.*, "A non-contacted height measurement method in two-dimensional space," *Sensors*, vol. 24, no. 21, Art. no. 6796, 2024, doi: [10.3390/s24216796](https://doi.org/10.3390/s24216796).
- [32] R. E. Caraka *et al.*, "Bayesian network analysis of spatial disparities in child nutrition status and health determinants across Indonesia," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 144, Art. no. 104936, 2025, doi: [10.1016/j.jag.2025.104936](https://doi.org/10.1016/j.jag.2025.104936).
- [33] S. Bakri, D. A. Puspaningrum, K. N. Berawi, and R. Zuraida, "Determinants of stunting severity among under-fives: Comparison of agricultural and nonagricultural households in South Lampung Regency, Indonesia," *Adv. Public Health*, vol. 2025, no. 1, Art. no. 7443296, 2025, doi: [10.1155/adph/7443296](https://doi.org/10.1155/adph/7443296).
- [34] W. R. Mgomezulu, P. Thangata, B. Mkandawire, and N. Amoah, "Advancing predictive analytics in child malnutrition: Machine, ensemble and deep learning models with balanced class distribution for early detection of stunting and wasting," *Hum. Nutr. Metab.*, vol. 42, Art. no. 200340, 2025, doi: [10.1016/j.hnm.2025.200340](https://doi.org/10.1016/j.hnm.2025.200340).
- [35] A. P. Joshi and B. V. Patel, "Data preprocessing: The techniques for preparing clean and quality data for data analytics process," *Orient. J. Comput. Sci. Technol.*, vol. 13, no. 2–3, pp. 78–81, 2021, doi: [10.13005/ojcs13.0203.03](https://doi.org/10.13005/ojcs13.0203.03).
- [36] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Front. Energy Res.*, vol. 9, Art. no. 652801, 2021, doi: [10.3389/fenrg.2021.652801](https://doi.org/10.3389/fenrg.2021.652801).
- [37] L. Wang and L. Liu, "Athlete body type recognition based on spectral co-clustering and CNN," *Int. J. Cloth. Sci. Technol.*, vol. 38, no. 3, pp. 389–414, 2026, doi: [10.1108/IJCST-04-2023-0055](https://doi.org/10.1108/IJCST-04-2023-0055).
- [38] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, 2022, doi: [10.1109/TNNLS.2021.3084827](https://doi.org/10.1109/TNNLS.2021.3084827).
- [39] C. Chen, N. A. Mat Isa, and X. Liu, "A review of convolutional neural network based methods for medical image classification," *Comput. Biol. Med.*, vol. 185, Art. no. 109507, 2025, doi: [10.1016/j.compbiomed.2024.109507](https://doi.org/10.1016/j.compbiomed.2024.109507).
- [40] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp. 92–108, 2022, doi: [10.1016/j.neucom.2022.06.111](https://doi.org/10.1016/j.neucom.2022.06.111).
- [41] Y. Li *et al.*, "A survey on dropout methods and experimental verification in recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 6595–6615, 2023, doi: [10.1109/TKDE.2022.3187013](https://doi.org/10.1109/TKDE.2022.3187013).
- [42] H. Khoshvaght, R. R. Permal, A. Razmjou, and M. Khiadani, "A critical review on selecting performance evaluation metrics for supervised machine learning models in wastewater quality prediction," *J. Environ. Chem. Eng.*, vol. 13, no. 6, Art. no. 119675, 2025, doi: [10.1016/j.jece.2025.119675](https://doi.org/10.1016/j.jece.2025.119675).
- [43] F. Roggio, B. Trovato, M. Sortino, and G. Musumeci, "A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: A

narrative review," *Heliyon*, vol. 10, no. 21, Art. no. e39977, 2024, doi: [10.1016/j.heliyon.2024.e39977](https://doi.org/10.1016/j.heliyon.2024.e39977).

- [44] A. Latreche, R. Kelaiaia, A. Chemori, and A. Kerboua, "Reliability and validity analysis of MediaPipe-based measurement system for some human rehabilitation motions," *Measurement*, vol. 214, Art. no. 112826, 2023, doi: [10.1016/j.measurement.2023.112826](https://doi.org/10.1016/j.measurement.2023.112826).
- [45] C. I. Allen Akselrud, "Random forest regression models in ecology: Accounting for messy biological data and producing predictions with uncertainty," *Fish. Res.*, vol. 280, Art. no. 107161, 2024, doi: [10.1016/j.fishres.2024.107161](https://doi.org/10.1016/j.fishres.2024.107161).
- [46] R. Dunne *et al.*, "Thresholding Gini variable importance with a single-trained random forest: An empirical Bayes approach," *Comput. Struct. Biotechnol. J.*, vol. 21, pp. 4354–4360, 2023, doi: [10.1016/j.csbj.2023.08.033](https://doi.org/10.1016/j.csbj.2023.08.033).
- [47] T. Thaventhiran *et al.*, "A digital health solution for child growth monitoring at home: Testing the accuracy of a novel 'GrowthMonitor' smartphone application to detect abnormal height and body mass indices," *Mayo Clin. Proc. Digit. Health*, vol. 1, no. 4, pp. 498–509, 2023, doi: [10.1016/j.mcpdig.2023.08.001](https://doi.org/10.1016/j.mcpdig.2023.08.001).
- [48] D. A. Rusakov, "A misadventure of the correlation coefficient," *Trends Neurosci.*, vol. 46, no. 2, pp. 94–96, 2023, doi: [10.1016/j.tins.2022.09.009](https://doi.org/10.1016/j.tins.2022.09.009).

Author Biography



Asmaul Husna, RS, is a doctoral (Ph.D.) candidate in the Vocational Education in Engineering Study Program at Makassar State University. She obtained her bachelor's degree in Educational Technology from Makassar State University in 2016 and continued

her master's studies in Image Processing at Hasanuddin University (UNHAS) in 2017. Her research interests focus on educational technology and artificial intelligence in the healthcare sector, particularly in the development of digital-based learning, the application of intelligent technologies to support education and healthcare services, and the enhancement of healthcare workforce competencies through innovative technologies that adapt to industrial advancements. She can be contacted at email: asmaul.husna@student.unm.ac.id



Purnamawati is a Full Professor in the Vocational Technology and Engineering Education Study Program at Makassar State University (UNM). She obtained her bachelor's degree in Electronics Engineering Education from IKIP Ujung Pandang in 1990, her master's degree in Vocational Technology and Engineering Education from Universitas Negeri Yogyakarta in 1996, and her doctoral degree in the same field from Universitas Negeri Yogyakarta in 2011. Her research interests encompass vocational education model development, metacognition-based learning, IoT-based learning media, teaching industry approaches, and the integration of digital technology in vocational and engineering education. She can be contacted at email: purnamawati@unm.ac.id



Hendra Jaya is a Full Professor in the Electronics Engineering Education Department, Faculty of Engineering, Makassar State University (UNM). He obtained his bachelor's degree in Electronics Engineering Education from Universitas Negeri Makassar in 2004 (cum laude), his master's degree in Electrical Engineering from Universitas Gadjah Mada in 2006, and his doctoral degree in Vocational Technology and Engineering Education from Universitas Negeri Yogyakarta in 2013 (cum laude). His research interests encompass virtual reality-based learning, remote laboratory systems, Internet of Things (IoT), artificial intelligence in education, and the development of simulation-based learning media for vocational education. He can be contacted at email: hendra.jaya@unm.ac.id



Anas Arfandi is an Associate Professor and Head of the Master's Program in Vocational Technology and Engineering Education at Makassar State University (UNM). He obtained his bachelor's degree in Building Engineering Education from Makassar State University in 2004, his master's degree in Vocational Technology and Engineering Education from Universitas Negeri Yogyakarta in 2007, and his doctoral degree in the same field from Universitas Negeri Yogyakarta in 2013. His research interests include vocational education curriculum development, higher-order thinking skills, work readiness, digital literacy, and the application of outcome-based education in technical and vocational education and training (TVET). He can be contacted at email: anas.arfandi@unm.ac.id