Received: 26 November 2021 | Revised: 24 March 2022 | Accepted: 30 March 2022

DOI: 10.1111/mcn.13364

ORIGINAL ARTICLE

Maternal & Child Nutrition WILEY

A novel screening tool to predict severe acute malnutrition through automated monitoring of weight-for-age growth curves

Sanja Nel[1,2,3] | Ute D. Feucht[2,3,4,5] | André L. Nel[6] | Piet J. Becker[7] | Friedeburg A. M. Wenhold[1,2,3]

[1]Department of Human Nutrition, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa

[2]Research Centre for Maternal, Fetal, Newborn & Child Health Care Strategies, University of Pretoria, Pretoria, South Africa

[3]Maternal and Infant Health Care Strategies Unit, South African Medical Research Council (SA MRC), Pretoria, South Africa

[4]Department of Paediatrics, University of Pretoria, Pretoria, South Africa

[5]District Clinical Specialist Team, Tshwane District Health Services, Gauteng Department of Health, Pretoria, South Africa

[6]Department of Mechanical Engineering, Faculty of Engineering and Built Environment, University of Johannesburg, Johannesburg, South Africa

[7]Research Office, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa

Correspondence
Sanja Nel, Department of Human Nutrition, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa.
Email: nel.sanja@gmail.com

Abstract

Weight-for-age (WFA) growth faltering often precedes severe acute malnutrition (SAM) in children, yet it is often missed during routine growth monitoring. Automated interpretation of WFA growth within electronic health records could expedite the identification of children at risk of SAM. This study aimed to develop an automated screening tool to predict SAM risk from WFA growth, and to determine its predictive ability compared with simple changes in weight or WFA *z*-score. To develop the screening tool, South African child growth experts (n = 30) rated SAM risk on 100 WFA growth curves, which were then used to train an artificial neural network (ANN) to assess SAM risk from consecutive WFA *z*-scores. The ANN was validated in 185 children under five (63 SAM cases; 122 controls) using diagnostic accuracy methodology. The ANN's performance was compared with that of changes in weight or WFA *z*-score. Even though experts' SAM risk ratings of the WFA growth curves differed considerably, the ANN achieved a sensitivity of 73.0% (95% confidence interval [CI]: 60.3; 83.4), specificity of 86.1% (95% CI: 78.6; 91.7) and receiver-operating characteristic curve area of 0.795 (95% CI: 0.732; 0.859) during validation with real cases, outperforming changes in weight or WFA *z*-scores. The ANN, as an automated screening tool, could markedly improve the identification of children at risk of SAM using routinely collected WFA growth information.

KEYWORDS

artificial intelligence; child growth monitoring; computer; electronic health records; failure to thrive; neural networks; nutrition screening; severe acute malnutrition

1 | INTRODUCTION

Children in South Africa (South African National Department of Health [NDOH], Statistics South Africa [SSA], South African Medical Research Council [SAMRC], 2017; Shisana et al., 2013) and other low-and middle-income countries (Fagbamigbe et al., 2020; Moyer et al., 2020) remain at risk of severe acute malnutrition (SAM). According to the World Health Organization (WHO), SAM is diagnosed in children aged 6–59 months with weight-for-length/height *z*-score < −3, mid-upper arm circumference < 11.5 cm, or bilateral oedema (WHO, United Nations Children's Fund [UNICEF], 2009).

The authors' personal observations and communication from colleagues working at public hospitals in Tshwane, Gauteng Province suggest that SAM is often preceded by growth faltering, characterized by slower-than-expected weight gain or even weight loss. Weight-for-age (WFA) growth is routinely monitored at South African primary healthcare (PHC) facilities as an early indicator of undernutrition or illness (South African NDOH, 2014, 2015). Unfortunately, effective growth monitoring is hindered by equipment unavailability, staff shortages and poor clinician proficiency in growth curve interpretation (Blaauw et al., 2017; Cloete et al., 2013; Kitenge & Govender, 2015). Consequently, even obvious WFA growth faltering is poorly identified, while weight-for-length/height and mid-upper arm circumference are rarely monitored at all (Blaauw et al., 2017; Kitenge & Govender, 2015).

Digitization of certain aspects of the South African Road-to-Health Booklet (RtHB) as an electronic clinic record has been considered by the South African Department of Health to address various surveillance problems in the growth monitoring programme. Electronic growth charts could integrate tools for automated growth curve interpretation, including an algorithm to identify WFA growth faltering. However, there is no clear, quantitative definition of growth faltering (Olsen, 2006; Roberfroid et al., 2005). Healthy growth maintains an approximately consistent WFA *z*-score over time, but some intraindividual variability is typical, and it is unclear what degree of deviation should be considered problematic (National Institutes for Health and Care Excellence, 2017; Roberfroid et al., 2005; WHO, 2012). In practice, clinicians' interpretation of a growth curve is largely qualitative, informed by clinical experience and context, making it difficult to quantify and reproduce.

Given the inherent complexity of growth curve interpretation, more sophisticated computational approaches—for example, machine learning—may prove more appropriate than simple rule-based algorithms for identifying growth faltering. Machine learning uses mathematical modelling to identify patterns in data and associate those patterns with specific outputs (Wiens & Shenoy, 2018). Artificial neural networks (ANNs) are a machine learning method that may be well suited to assessing growth curves. Programming or 'training' an ANN is analogous to human learning: the ANN is presented with a large data set of inputs and associated outputs, from which it develops a classification system to associate any given input with the appropriate output (Bishop, 2006; Engelbrecht, 2007; Wiens & Shenoy, 2018). For this study, the ANN was trained using growth curves (presented as a series of consecutive WFA *z*-scores) as input, with associated classifications of SAM risk as output.

Key messages

- Growth faltering preceding severe acute malnutrition (SAM) is poorly identified during routine growth monitoring, and opportunities for preventative interventions are missed.
- Consistently identifying growth faltering is challenging, even for experts in child growth. Digitized growth monitoring tools incorporating automated pattern recognition by an artificial neural network (ANN) can assist clinicians in identifying growth faltering.
- In this study, an ANN was trained to identify children at risk of SAM based on weight-for-age growth curves, with promising results. With further refinement, the ANN could greatly improve the identification of children with growth faltering and facilitate earlier interventions to prevent SAM.

The first phase of this study aimed to develop an ANN for predicting SAM risk from WFA growth curves, while the second phase evaluated the predictive validity of the ANN using real-life clinical cases, and compared it with simple mathematical indicators based on changes in weight or WFA *z*-score.

2 | MATERIALS AND METHODS

2.1 | Development of the ANN

Training data for the ANN were collected through a survey of South African child growth experts. Experts were recruited purposively by contacting the Departments of Human Nutrition/Dietetics, Public Health and Paediatrics at 11 South African universities and the Dietetics Departments of 10 tertiary/academic public hospitals. Participants were also asked to recommend colleagues for recruitment.

To obtain input data for training the ANN, 100 WFA growth curves were compiled from a selection of local and international growth monitoring training materials (Chopra et al., 2000; Savage & Copeman, 2006; Savage King & Burgess, 1993; Uganda Ministry of Health, 2003; WHO, 2012), plus self-developed curves for inadequately represented clinical scenarios. For each growth curve, WFA *z*-scores and digital images were generated using WHO Anthro software v3.2.2 (2011; WHO) and Microsoft Excel (see Supporting Information: Appendix). Output data for training the ANN were obtained from respondents' classification of each WFA growth curve as representing low, medium or high risk of SAM (the operational meanings of which were clearly defined in the survey introduction).

The 100 charts were divided into two comparable surveys of 50 charts each, to lessen the respondent time burden. Experts were

randomly assigned to receive either Survey A or Survey B. Each electronic survey (developed using Google Forms) consisted of instructions, informed consent, biographical questions and 50 growth chart images presented in random order. Three charts were repeated per survey to assess intrarater reliability. Interrater reliability (agreement) was investigated for each survey in terms of the number of risk categories that were selected for each chart, and the percentage of experts who agreed on each rating. Descriptive data for the two groups of expert respondents were compared using Fisher's exact test.

The WFA *z*-scores (input) and respondents' SAM risk classification (output) for each curve were used to train the ANN (Mathematica v12.0, 2019; Wolfram Research). Various computational approaches were attempted, including *K*-nearest neighbour, random forest and multilayer perceptron approaches. For each training attempt, 10% of the charts were randomly removed before training and used for posttraining testing. The ANN with the highest posttraining test score was then validated.

The presence of oedema, although an important factor in the diagnosis of SAM, was not explicitly included in the training of the ANN. Instead, a reminder check for oedema will be built into the final application as a separate function and a positive response (i.e., oedema is present) will result in a recommendation to refer the child to a higher level of care regardless of weight.

2.2 | ANN validation

The predictive validity of the ANN was evaluated using a diagnostic accuracy assessment methodology with a case–control design. Children under five with SAM (cases) were recruited from two public-sector hospitals in Tshwane District, Gauteng Province, South Africa. SAM was diagnosed using WHO criteria, as the presence of any one of a weight-for-length/height *z*-score < −3, mid-upper arm circumference < 11.5 cm or bilateral oedema (5). Controls—children under five without SAM—were recruited as a consecutive convenience sample at three PHC clinics in these hospitals' referral areas (July 2018–2019). Availability of a patient-held RtHB with at least three recorded weights was required for inclusion. Preterm-born children and children with medical/congenital conditions requiring disease-specific growth references were excluded.

An *a priori* power calculation (nQuery Advanced v8.1.0.0. 2017; Statistical Solutions Ltd.) indicated that 63 cases and 126 controls would be sufficient to reject H_0: Sensitivity = 75% in favour of H_a: Sensitivity ≥ 90%.

Sociodemographic, medical and birth information were collected from the RtHB and parent/caregiver interview. All past weight measurements (and associated dates) were obtained from the RtHB. At recruitment, weight was measured by clinic nurses during routine growth monitoring (for controls) or by hospital dietitians at admission (for cases) using the facilities' electronic scales and recorded at 0.1 kg. Facility weights were used unaltered to reflect a real-world scenario. Length/height was measured and recorded at 0.1 cm. Cases were measured at admission by hospital dietitians using hospital equipment, while controls were measured at recruitment by the investigator (S. N.), using a rigid wooden length board (ShorrBoard Portable Height/Length Measuring Board; Weight and Measure LLC) for recumbent length (≤2 years) and a rigid, free-standing stadiometer (Leicester Height Meter; Seca) for standing height (>2 years).

2.2.1 | Data preparation and analysis

Growth parameters were converted to *z*-scores using WHO Anthro v3.2.2 (2011; WHO). Data analysis was performed using Stata version 15.1 (2017, StataCorp LLC). Significance was set at 5% using two-sided probabilities. Descriptive data for the case and control groups were compared using Student's *t* test (normally distributed continuous variables), two-sample Wilcoxon's rank-sum (Mann–Whitney) test (nonnormally distributed continuous variables) and Fisher's exact test (categorical variables).

Diagnostic accuracy calculations (as per STARD 2015; Cohen et al., 2016) were performed, comparing the risk of SAM as exposure to the presence/absence of SAM as the outcome. To define the risk of SAM for the exposure variable, each child was classified as being 'at risk' or 'not at risk' of SAM based on their WFA growth curve, using all weights recorded in the RtHB preceding the weight that was used to define the presence/absence of SAM. Two approaches were used:

(a) ANN: The ANN classified each child's consecutive WFA *z*-scores as representing low, medium or high risk of SAM. To facilitate diagnostic testing, it was decided *a priori* to combine the medium- and high-risk groups into a single 'at risk' class, with the low-risk class considered as 'not at risk'.
(b) Simpler mathematical indicators: The difference between the last two recorded weights (preceding the final weight measurement that was used to define the outcome) and their WFA *z*-scores were obtained by subtracting the earlier value from the latter. Six cut-offs were evaluated for indicating a child as 'at risk' of SAM: weight loss (difference < 0), weight stagnation/loss (difference ≤ 0), any decrease in WFA *z*-score and decreases of more than 0.33, 0.50 and 0.67 WFA *z*-scores, respectively.

The primary analyses focused on sensitivity and specificity, as (unlike predictive values) they are unaffected by the high proportion of SAM cases in the sample. The area under the receiver-operating characteristic curve (ROC-AUC) was calculated as a measure of the balance between sensitivity and specificity for each indicator of SAM risk. To explore the effect of SAM prevalence on the performance of the ANN, the positive and negative predictive values (PPV and NPV) were recalculated using a hypothetical sample of 1000 children with four different SAM prevalence rates: an estimated average under-5 SAM prevalence of 1%, based on the prevalence of severe wasting described in large-scale South African nutrition surveys (South African NDOH, SSA, SAMRC, 2017; Shisana et al., 2013) as well as 0.5%, 3.0%, 5.0% and 10% to allow for variation in prevalence.

3 | RESULTS

3.1 | ANN development

Expert survey responses: From 69 experts contacted, 30 responses were received (Survey A: 13 responses; Survey B: 17 responses; response rate = 43.5%). The professional profile of the respondents (Table 1) evinces their expertise in child growth and nutrition.

Comparison of expert responses revealed substantial inter-rater disagreement, as shown in Table 2. All respondents agreed on the risk classification of only 10/100 charts, while 41/100 charts were simultaneously classified as low, medium and high risk by different respondents. For 15/100 charts, <50% of the experts agreed on the most-selected rating. Calculation of Cohen's κ confirmed the lack of agreement among the respondents, with κ values for the different risk categories ranging from 0.16 to 0.59 (Table S1) Interrater agreement varied with the growth pattern represented by the charts, being highest for charts representing healthy weight gain and lowest for charts of low-birth-weight children (Table S2).

TABLE 1 Professional profile of respondents.

	Total (N = 30), n (%)	Survey A (N = 13), n (%)	Survey B (N = 17), n (%)	p Value[d] (difference)
Place of work[a]				
University	12 (40.0)	6 (46.2)	6 (35.3)	0.360
Public health facility	17 (56.7)	6 (46.2)	11 (64.7)	
Nongovernmental	1 (3.3)	1 (7.7)	0	
Nature of current and recent work (relating to child growth and nutrition)				
Teaching at a tertiary institution	15 (50.0)	6 (46.2)	9 (52.9)	1.000
In-service training of healthcare providers	9 (30.0)	4 (30.8)	5 (29.4)	1.000
Research	9 (30.0)	4 (30.8)	5 (29.4)	1.000
Policy development	7 (23.3)	3 (23.1)	4 (23.5)	1.000
Clinical work: Primary healthcare level	2 (6.7)	1 (7.7)	1 (5.9)	1.000
Clinical work: Hospital level	10 (33.3)	3 (23.1)	7 (41.2)	0.440
Health promotion at the community level	2 (6.7)	2 (15.4)	0	0.179
Other	1 (3.3)	1 (7.7)	0	0.433
Profession[b]				
Medical practitioner (doctor)	9 (30.0)	3 (23.1)	6 (35.3)	0.691
Dietitian/nutritionist	21 (70.0)	10 (76.9)	11 (64.7)	
Highest qualification[c]				
Bachelor's degree (4+ years)	2 (6.7)	2 (15.4)	0	0.150
Bachelor's plus postgraduate diploma	5 (16.7)	1 (7.7)	4 (23.5)	
Master's degree	17 (56.7)	6 (46.2)	11 (64.7)	
PhD degree or equivalent	6 (20.0)	4 (30.8)	2 (11.8)	
Years' experience working in child health and nutrition				
0–3	1 (3.3)	0	1 (5.9)	0.974
4–7	8 (26.7)	3 (23.1)	5 (29.4)	
8–12	7 (23.3)	4 (30.8)	3 (17.6)	
13–20	5 (16.7)	2 (15.4)	3 (17.6)	
>20	9 (30.0)	4 (30.8)	5 (29.4)	

[a]No respondent selected 'Private sector', 'Research entity' or 'Other'.

[b]No respondent selected 'Nurse'.

[c]No respondent selected 'Diploma' or 'PhD plus postdoctoral'.

[d]Fisher's exact test.

TABLE 2 Expert agreement on risk of severe acute malnutrition, represented by weight-for-age growth charts.

	Survey A (13 respondents; *N* = 50 charts), *n* (%) of charts[a]	Survey B (17 respondents; *N* = 50 charts), *n* (%) of charts[a]	Total (30 respondents; *N* = 100 charts), *n* (%) of charts[a]
Number of risk classes selected by different experts for the same growth chart			
1	6 (12.0)	4 (8.0)	10 (10.0)
2	28 (56.0)	21 (42.0)	49 (49.0)
3	16 (32.0)	25 (50.0)	41 (41.0)
% of experts agreeing on the most-selected risk class per chart			
100	6 (12.0)	4 (8.0)	10 (10.0)
75–<100	17 (34.0)	13 (26.0)	30 (30.0)
50–<75	22 (44.0)	23 (46.0)	45 (45.0)
<50	5 (10.0)	10 (20.0)	15 (15.0)

[a]Each survey consisted of 50 charts. Each chart was rated as low risk/medium risk/high risk.

Intrarater (test–retest) reliability varied. In both surveys, the chart representing a clear high- or low-risk growth pattern was classified identically both times by all respondents. The four charts with more ambiguous or complex growth patterns were rated identically both times by 9/13 (69.2%), 10/13 (76.9%), 12/17 (70.6%) and 14/17 (82.4%) of respondents, respectively. McNemar's test for symmetry revealed no directional bias ($p = 0.223–0.606$).

3.1.1 | ANN training

The ANN was initially trained using each individual expert's classification of each chart as a separate input, yielding a total of 1500 growth charts (1350 for training, 150 for testing). However, the inconsistency in the expert responses prevented the ANN from converging on an effective solution, although the multilayer perceptron approach with backpropagation learning (training accuracy = 54.5%) was found to be superior to *K*-nearest neighbour (training accuracy = 18.2%) and random forest (training accuracy = 35.7%) approaches.

To improve training accuracy, the experts' responses were combined to give a single SAM risk classification for each chart. Each individual response was assigned a numeric value (low risk = −1, moderate risk = 0 and high risk = +1) and the mean of all responses was calculated for each chart. The mean value was then assigned a risk category (low risk = −1 to −0.66, moderate risk = −0.67 to +0.66 and high risk = +0.67 to +1). This yielded a training data set of 100 growth charts (90 for training, 10 for testing). The ANN trained with this data set (using the multilayer perceptron approach) achieved sufficient posttraining accuracy (73.3%) to justify validation with real clinical data.

3.2 | ANN validation

3.2.1 | Sample description

A total of 63 SAM cases and 126 controls were recruited, but four controls were excluded due to incomplete growth information. The sample is described in Table 3.

Cases and controls differed significantly in their age, who they lived with, HIV status, immunization status and number of comorbidities. Further analyses revealed significant differences in patterns of comorbidities: mild upper respiratory tract infections were the most common comorbidity among controls ($n = 12$ controls (9.8%), 0 cases; $p = 0.009$), while cases had higher prevalence of acute gastroenteritis ($n = 15$ cases (23.8%), 1 control (0.8%); $p < 0.001$) and bronchopneumonia ($n = 3$ cases (4.8%), 0 controls; $p = 0.038$). Oedema was reported in 25 (39.7%) of the cases.

3.2.2 | Diagnostic accuracy

The diagnostic accuracy of the ANN and simpler mathematical indicators of SAM risk are described in Table 4. Any decrease in the WFA *z*-score had the highest sensitivity (77.8%), but the specificity was poor (49.3%). Weight loss had the highest specificity (91.0%), but at the cost of sensitivity (44.4%). The ANN had the highest ROC-AUC (0.795), indicating a good balance of high sensitivity (73.0%) and specificity (86.1%). There was minimal overlap between the 95% confidence intervals of the ANN and the weight-based indicators, and no overlap with the WFA *z*-score-based indicators, affirming its superior predictive ability. Due to the significant difference in the ages of the cases and controls, the diagnostic accuracy analysis was repeated with only children aged 0–24 or 6–24 months. This had minimal effect on the diagnostic performance in the 0–24 months group ($n = 168$; sensitivity = 72.6%, specificity = 86.8%, ROC-AUC = 0.797), although specificity was somewhat decreased for the 6–24 months group ($n = 120$; sensitivity = 73.3%, specificity = 81.7%, ROC-AUC = 0.775).

Table 5 shows the positive and negative predictive values for the ANN at various SAM prevalence rates, with 34% representing the artificially high prevalence rate in the validation sample. As SAM prevalence decreases, the PPV declines rapidly (from 21.7% at 5.0% prevalence to 2.6% at 0.5% prevalence), while the NPV increases (from 98.4% at 5.0% prevalence to 99.8% at 0.5% prevalence).

4 | DISCUSSION

This study highlighted two important findings: first, the interpretation of WFA growth curves may be more difficult than is commonly believed (as evidenced by the disagreement among the surveyed experts). and second, automated interpretation of WFA growth curves shows promise in improving the identification of children at risk of SAM.

TABLE 3 Description of the validation sample.

Characteristic	Controls (*N* = 122), *n* (%)[a]	Cases (N = 63), *n* (%)[a]	*p* Value[b] (difference)
Age category (months)			
<6	46 (37.7)	2 (3.2)	<0.001
6–<12	32 (26.2)	24 (38.1)	
12–<24	28 (23.0)	36 (57.1)	
≥24	16 (13.1)	1 (1.6)	
Sex			
Male	58 (47.5)	40 (63.5)	0.060
Female	62 (50.8)	23 (36.5)	
Lives with			
Both parents	84 (68.9)	18 (28.6)	<0.001
Mother only	34 (27.9)	40 (63.5)	
Other	1 (0.8)	4 (6.3)	
Race			
Black African	116 (95.1)	60 (95.2)	0.800
Other	5 (4.9)	0	
Neonatal feeding (first month of life)			
Exclusive breastfeeding	99 (81.1)	47 (74.6)	0.760
Exclusive formula feeding	15 (12.3)	9 (14.3)	
Mixed feeding	4 (3.3)	1 (1.6)	
HIV status			
Exposure unknown	3 (2.5)	1 (1.6)	<0.001
Unexposed	85 (69.7)	28 (44.4)	
Exposed, HIV infection status unknown	6 (4.9)	2 (3.2)	
Exposed, confirmed HIV uninfected	25 (20.5)	19 (30.2)	
HIV-positive, on HAART	1 (0.8)	8 (12.7)	
HIV-positive, not on HAART	0	3 (4.8)	
Immunizations			
Up to date	117 (95.9)	46 (73.0)	<0.001
Not up to date	5 (4.1)	16 (25.4)	
Current illness (number of acute comorbidities)			
0	108 (88.5)	32 (50.8)	<0.001
1	12 (9.8)	27 (42.9)	
2 or more	2 (1.6)	4 (6.4)	
Age (months)			
Mean (SD)	12.5 (12.4)	13.7 (6.5)	-[c]
Median (IQR)	7.8 (4.1–17.7)	12.7 (9.7–16.9)	<0.001[d]
Birthweight (kg)			
Mean (SD)	3.2 (0.4)	3.1 (0.4)	0.035[e]
Median (IQR)	3.2 (2.9–3.4)	3.0 (2.9–3.3)	0.030[d]

TABLE 3 (Continued)

Characteristic	Controls (*N* = 122), *n* (%)[a]	Cases (N = 63), *n* (%)[a]	*p* Value[b] (difference)
Birth length (cm)			
Mean (SD)	50.5 (4.0)	50.1 (0.6)	0.566[e]
Median (IQR)	51.0 (49.0–52.0)	50.0 (49.0–52.0)	0.233[d]

Abbreviations: HAART, highly active antiretroviral therapy; HIV, human immunodeficiency virus; IQR, interquartile range; NA, not applicable; SD, standard deviation.

[a]Where %s in a category do not add up to 100, data were missing for some respondents.

[b]Fisher's exact test used, unless specified otherwise.

[c]Analysis of difference between means inappropriate (nonnormal distribution).

[d]Two-sample Wilcoxon's rank-sum (Mann–Whitney) test.

[e]Student's *t* test.

TABLE 4 Predictive validity of WFA growth faltering-related indicators of SAM, expressed in terms of diagnostic accuracy parameters.

Indicator of SAM risk	*n* classified 'at risk'		Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	ROC-AUC area (95% CI)
	Cases (*n* = 63)	Controls (*n* = 122)			
Artificial neural network rating	46	17	73.0 (60.3; 83.4)	86.1 (78.6; 91.7)	0.795 (0.732; 0.859)
Any weight loss	28	11	44.4 (31.9; 57.5)	91.0 (84.4; 95.4)	0.677 (0.610; 0.744)
Weight stagnation/loss	30	17	47.6 (34.9; 60.6)	86.1 (78.6; 91.7)	0.668 (0.599; 0.738)
Any decrease in WFA *z*-score	49	62	77.8 (65.5; 87.3)	49.2 (40.0; 58.4)	0.635 (0.567; 0.703)
WFA *z*-score decrease >0.33	38	39	60.3 (47.2; 72.4)	68.0 (59.0; 76.2)	0.642 (0.568; 0.715)
WFA *z*-score decrease >0.50	32	29	50.8 (37.9; 63.6)	76.2 (67.7; 83.5)	0.635 (0.562; 0.708)
WFA *z*-score decrease >0.67	26	15	41.3 (29.0; 54.4)	87.7 (80.5; 93.0)	0.645 (0.577; 0.713)

Abbreviations: CI, confidence interval; ROC-AUC, area under the receiver operating characteristic curve; SAM, severe acute malnutrition; WFA, weight for age.

TABLE 5 Effect of changes in SAM prevalence on the positive and negative predictive values of the ANN

Prevalence (%)	*n* (SAM)	*n* (no SAM)	True positives (*a*)	False positives (*b*)	True negatives (*c*)	False negatives (*d*)	PPV (%) (*a*/[*a* + *b*])	NPV (%) (*c*/[*c* + *d*])
34.0	340	660	248.2	91.7	568.3	91.8	73.0	86.1
10.0	100	900	73	125,1	774,9	27	36.9	96.6
5.0	50	950	36.5	132.1	818.0	13.5	21.7	98.4
3.0	30	970	21.9	134.8	835.2	8.1	14.0	99.0
1.0	10	990	7.3	137.6	852.4	2.7	5.0	99.7
0.5	5	995	3.7	138.3	856.7	1.4	2.6	99.8

Note: Calculated for a hypothetical sample of *N* = 1000, using results from the ANN with sensitivity = 73.0% and specificity = 86.1%.

Abbreviations: ANN, artificial neural network; NPV, negative predictive value; PPV, positive predictive value; SAM, severe acute malnutrition.

Calculations: *n*(SAM) = *N* × prevalence; *n*(no SAM) = N - *n*(SAM); true positives = *n*(SAM) × sensitivity; false positives = *n*(no SAM) − true negatives; true negatives = *n*(no SAM) × specificity; false negatives = *n*(SAM) − true positives.

Due to the qualitative nature of growth curve interpretation, differences of opinion between clinicians are inevitable. However, in this survey disagreement predominated to an unexpected extent. It is well documented that frontline PHC health workers struggle with interpreting growth charts (Blaauw et al., 2017; Cloete et al., 2013; Kitenge & Govender, 2015), but the results of this survey suggest that the same may be true even for experts in the field. The lack of a 'gold standard' definition for growth faltering may be partly to blame, as it negatively affects reproducibility (i.e., interrater reliability) (Olsen, 2006; Roberfroid et al., 2005). This further highlights the

potential value of an automated growth monitoring tool to assist with clinical decision making at the PHC level.

Despite the inconsistencies in the training data, the ANN was able to identify children at risk of SAM based on WFA *z*-scores: in the validation sample, the ANN identified 46/63 cases (73.0%) as being at risk of SAM before the onset of SAM, while wrongly classifying only 17/122 (13.9%) of the controls as at risk of SAM. The ROC-AUC (0.795) affirms the ANN's superior ability to distinguish between children with and without SAM, compared with the other indicators assessed. WFA growth as a predictor of SAM has not been widely studied, making it difficult to determine whether the ANN could outperform well-trained clinicians. However, the ANN did outperform the status quo: none of the 63 included SAM cases had been identified as being at risk during routine growth monitoring, but the ANN was able to identify the risk of SAM at the same time point in 46 of these cases.

There were a number of differences between the validation sample cases and controls. These factors may be related to the aetiology of SAM, including living with the mother only (which may affect household income and food security), higher rates of HIV infection and other acute comorbidities in cases, and lower rates of up-to-date immunization, which may suggest poor attendance of routine growth monitoring. On the one hand, this affects the comparability of the cases and controls, but on the other hand, it provides important insights into additional factors that may distinguish children with SAM from those without SAM. Selectively sampling SAM cases to be comparable to controls, particularly with regard to HIV infection and comorbidities, would exclude a large and clinically significant portion of children with SAM.

This study evaluated the sensitivity and specificity of the ANN in a sample with a predetermined case:control ratio. However, PPV and NPV —that is, the degree to which a positive/negative test result predicts the true presence/absence of the target condition—are affected by the population prevalence of the target condition (Gleason et al., 2010). The variation in the PPV and NPV of the ANN according to SAM prevalence (Table 5) suggests that the screening tool would be of greater practical value in populations with a higher prevalence of SAM, as the PPV is higher. A low PPV implies a high number of false positives—many children are identified as being at risk of SAM when they are not truly at risk. The management of false positives carries resource costs (primarily nursing time), which may place unnecessary strain on an already overburdened growth monitoring programme (Blaauw et al., 2017; Kitenge & Govender, 2015). Thus, the value of the ANN as a screening tool in PHC depends on both scientific and practical considerations, including local SAM prevalence and resource availability.

It must be acknowledged that WFA growth is an imperfect indicator for predicting SAM, which is diagnosed by entirely different criteria, namely, weight-for-length/height, mid-upper arm circumference and/or oedema (WHO, UNICEF, 2009). Anecdotal reports from local clinicians suggest that WFA growth faltering commonly precedes SAM. There is a paucity of published research on the WFA growth in children with SAM. In this study, 30/63 (47.6%) of SAM cases had a history of weight stagnation/loss before the SAM diagnosis, supporting the notion that WFA growth faltering is a reasonably common antecedent to SAM.

As the aetiology of SAM is often multifactorial, other socio-demographic and/or medical information may be useful to assess SAM risk (Rieger & Trommlerová, 2016). For example, a study conducted in the Democratic Republic of the Congo identified nine indicators (birthweight, diarrhoea, number of daily meals, duration of breastfeeding, age at complementary food introduction, maternal age, parity, family SAM history and number of under-5 children) as important predictors of SAM risk (Mukuku et al., 2019). A scoring system incorporating these factors could predict SAM risk with 93.5% sensitivity and 93.1% specificity, far outperforming this study's WFA-based indicators. Similarly, a study using machine learning to predict underweight (i.e., low WFA) in Bangladeshi children developed a model (using child age, maternal education level, wealth index, place of residence, maternal BMI and birth interval) that achieved a sensitivity of up to 94.66% (albeit with lower specificity of 69.76%) (Talukder & Ahammed, 2020). Taken together, these studies suggest that WFA growth may not be the single most important factor to consider when evaluating SAM risk. Nonetheless, weight is routinely measured, and the opportunity for identifying children at risk of SAM should not be wasted. Automated assessment of WFA growth could be a useful first step to identify children who need more in-depth assessment, which may include further anthropometric measurements (e.g., weight-for-length/height and mid-upper arm circumference) and other aspects of nutrition assessment.

Some limitations of this study must be acknowledged. First, the expert survey recruited a relatively small sample, yielding a small training data set that likely had a negative impact on ANN training. Second, due to the purposive sampling procedure, the experts cannot be considered representative of all South African child health and nutrition experts. This limits the generalizability of the findings regarding interrater agreement. Similarly, the sampling design for the validation study limits the generalizability of the findings to an unselected population of children, a fact underscored by the effect of SAM prevalence on the PPV and NPV of the ANN.

The exclusion of children born preterm is an important limitation, as these children may be at increased risk of SAM. This may also have affected the age distribution of the SAM cases, reducing the number of SAM cases below the age of 6 months. Future validation of the ANN with preterm-born children will be important to address this gap. The validation sample also contained few children over the age of 2 years (particularly cases); therefore, further validation with SAM cases aged 2–5 years would be needed before the ANN can be used with confidence in this age range. Nonetheless, the highest incidence of SAM is seen in children under the age of 2 years, and thus the ANN has been validated in the most important target audience.

This study pioneers a novel approach to identify children at risk of SAM, using WFA data that are routinely collected at PHC in most low- and middle-income countries (De Onis et al., 2012). However, some questions remain to be explored by future research. The value of WFA as a predictor of SAM, compared with other anthropometric and nonanthropometric indicators should be assessed, and the best, population-specific combination of indicators to predict SAM risk should

be identified. The predictive value of serial measurements of MUAC, in particular, deserves investigation: MUAC is one of the diagnostic indicators for SAM, and a declining MUAC may be just as valuable as weight loss in identifying children at risk of SAM. Unfortunately, MUAC is not routinely measured at most PHC facilities, which precluded its inclusion in the current study. In addition to growth parameters, other sociodemographic, health-related and dietary factors should be investigated as predictors of SAM. Additionally, appropriate responses to the identification of 'at-risk' children should be investigated, including further assessment (measurement of length/height and MUAC, dietary assessment), nutrition intervention (counselling, education and/or supplementation), appropriate referral (to nutrition services or higher levels of care) and more rapid follow-up. The safety and effectiveness of these interventions as well as their feasibility within the healthcare system need to be evaluated, particularly in light of the higher numbers of false positives that will be seen in areas with low SAM prevalence.

Future steps for improving the ANN include increasing the size of the training data set, ideally including real cases with a known outcome in the training data. Furthermore, before implementation of the ANN in any real-world setting can be considered, it must be validated in children born preterm, as they were excluded from this study. Finally, the programme or application in which the ANN will be embedded needs to be developed, with the inclusion of other relevant factors related to the aetiology and/or diagnosis of SAM, such as examination for oedema. The ANN is an important and ground-breaking first step, yet much work remains to be done to develop a fully functional clinical tool for the identification of children at risk of SAM.

5 | CONCLUSIONS

Automated evaluation of WFA growth curves by an ANN shows promise in identifying children at risk of SAM. Further refinements, such as incorporating assessment for oedema and the presence of additional SAM risk factors, could further improve the predictive validity of the ANN. Using automated growth curve assessment in routine growth monitoring, as opposed to relying on health workers' judgement alone, could improve the early identification of children at risk of SAM, thus facilitating targeted and timely interventions to prevent SAM.

AUTHOR CONTRIBUTIONS

Sanja Nel was the principal investigator in charge of planning and executing the research project. Friedeburg A. M. Wenhold and Ute D. Feucht guided the design and execution of the research. André L. Nel and Sanja Nel developed the survey instrument. André L. Nel trained the neural network and used it to evaluate all the participant data for severe acute malnutrition risk. Piet J. Becker performed all statistical analyses. Sanja Nel, Friedeburg A. M. Wenhold, and Ute D. Feucht wrote the paper.

ACKNOWLEDGEMENTS

The authors extend their thanks to the management and staff of the affected clinics and hospitals for allowing data collection at the facilities, with special thanks to the dietitians and research assistants at the hospitals for assisting with data collection.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data are available from the authors on request.

ETHICS STATEMENT

Ethical approval was obtained (University of Pretoria Faculty of Health Sciences Research Ethics Committee: protocol 468-2017), and permission to collect data was obtained from health authorities and the health institutions involved. Informed consent was obtained from the parent/caregiver for all participants. Feedback was given to all institutions and stakeholders, including experts who participated in the initial survey, institutions where data were collected, and the Gauteng Department of Health.

ORCID

Sanja Nel http://orcid.org/0000-0003-3175-7340
Ute D. Feucht http://orcid.org/0000-0002-6339-1040
André L. Nel http://orcid.org/0000-0002-2353-4377
Piet J. Becker http://orcid.org/0000-0002-9384-6472
Friedeburg A. M. Wenhold https://orcid.org/0000-0003-1140-5065

REFERENCES

Bishop, C. (2006). *Pattern recognition and machine learning*. Springer.

Blaauw, R., Daniels, L., du Plessis, L. M., Koen, N., Koornhof, H. E., Marais, M. L., vanNiekerk, E., & Visser, J. (2017). Assessing the utilisation of a child health monitoring tool. *South African Journal of Child Health*, *11*(4), 174–179. https://doi.org/10.7196/sajch.2017.v11i4.1326

Chopra, M., Gachuhi, D., Piwoz, E., Puoane, T., Sanders, D., & Swart, R. (2000). *Integrating nutrition into health and development programs: A training guide*. Health Systems Trust. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.473.5496&rep=rep1&type=pdf

Cloete, I., Daniels, L., Jordaan, J., Derbyshire, C., Volmink, L., & Schubl, C. (2013). Knowledge and perceptions of nursing staff on the new Road to Health Booklet growth charts in primary healthcare clinics in the Tygerberg Subdistrict of the Cape Town Metropole District. *South African Journal of Clinical Nutrition*, *26*, 141–146. https://doi.org/10.1080/16070658.2013.11734458

Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., Irwig, L., Levine, D., Reitsma, J. B., deVet, H. C., & Bossuyt, P. M. (2016). STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *BMJ Open*, *6*, e012799. https://doi.org/10.1136/bmjopen-2016-012799

DeOnis, M., Onyango, A. W., Borghi, E., Siyam, A., Blössner, M., & Lutter, C. (2012). Worldwide implementation of the WHO Child Growth Standards. *Public Health Nutrition*, *15*, 1603–1610. https://doi.org/10.1017/S136898001200105X

Engelbrecht, A. P. (2007). *Computational intelligence: An introduction* (2nd ed.). Wiley.

Fagbamigbe, A. F., Kandala, N.-B., & Uthman, O. A. (2020). Severe acute malnutrition among under-5 children in low- and middle-income countries: A hierarchical analysis of associated risk factors. *Nutrition*, *75-76*, 110768. https://doi.org/10.1016/j.nut.2020.110768

Gleason, P., Harris, J., Sheean, P., Boushey, C., & Breummer, B. (2010). Publishing nutrition research: Validity, reliability, and diagnostic test assessment in nutrition-related research. *Journal of the American Dietetic Association*, *110*, 409–419. https://doi.org/10.1016/j.jada.2009.11.022

Kitenge, G., & Govender, I. (2015). Nurses' monitoring of the Road to Health Chart at primary healthcare level in Makhado, Limpopo Province. *South African Family Practice*, *55*, 275–280. https://doi.org/10.1080/20786204.2013.10874350

Moyer, J. D., Bohl, D. K., Petry, C., Scott, A., Solórzano, J. R., & Kuhn, R. (2020). The persistent global burden of severe acute malnutrition: Cross-country estimates, models and forecasts. *Global Transitions*, *2*, 167–179. https://doi.org/10.1016/j.glt.2020.07.004

Mukuku, O., Mutombo, A. M., Kamona, L. K., Lubala, T. K., Mawaw, P. M., Aloni, M. N., Wembonyama, S. O., & Luboya, O. N. (2019). Predictive model for the risk of severe acute malnutrition in children. *Journal of Nutrition and Metabolism*, *2019*, 4740825.

National Institutes for Health and Care Excellence (NICE). (2017). *Faltering growth—Recognition and management (clinical guideline NG75): Methods, evidence and recommendations*. Retrieved August 2, 2021, from https://www.nice.org.uk/guidance/NG75

Olsen, E. M. (2006). Failure to thrive: Still a problem of definition. *Clinical Pediatrics*, *45*, 1–6. https://doi.org/10.1177/000992280604500101

Rieger, M., & Trommlerová, S. (2016). Age-specific correlates of child growth. *Demography*, *53*, 241–267. https://doi.org/10.1007/s13524-015-0449-3

Roberfroid, D., Kolsteren, P., Hoeree, T., & Maire, B. (2005). Do growth monitoring and promotion programs answer the performance criteria of a screening program? A critical analysis based on a systematic review. *Tropical Medicine and International Health*, *10*, 1121–1133. https://doi.org/10.1111/j.1365-3156.2005.01498.x

Savage, F., & Copeman, J. (2006) Growth lines of undernourished children: Slides extracted from TALC slide set 'Undernutrition in young children (PED)'. In: UNICEF, TALC. *Community nutrition: Nutrition materials for healthcare workers in resource-poor areas* [CD-ROM]. Health Books International. https://healthbooksinternational.org/product/community-nutrition/

Savage King, F., & Burgess, A. (1993). *Nutrition for developing countries* (2nd ed.). Oxford University Press.

Shisana, O., Labadarios, D., Rehle, T., Simbayi, L., Zuma, K., Dhansay, A., & Faber, M. (2013). *South African National Health and Nutrition Examination Survey (SANHANES-1)*. HSRC Press. Retrieved August 2, 2021, from http://www.hsrc.ac.za/en/research-outputs/view/6493

South African National Department of Health (NDOH). (2014). *Integrated Management of Childhood Illness (IMCI) booklet*. National Department of Health. Retrieved August 2, 2021, from https://fundisa.ac.za/wp-content/uploads/2015/10/IMCI-booklet-2014.pdf

South African National Department of Health (NDOH). (2015). *Integrated management of children with acute malnutrition in South Africa: Operational guidelines*. National Department of Health. Retrieved August 2, 2021, from https://profmoosa.com/wp-content/uploads/2019/09/guidelines_acute_malnutrition_south_africa_2015.pdf

South African National Department of Health (NDOH), Statistics South Africa (SSA), South African Medical Research Council(SAMRC). (2017). *South Africa Demographic and Health Survey 2016: Key indicators*. National Department of Health. Retrieved August 2, 2021, from http://www.statssa.gov.za/?page_id=1854&PPN=Report%2003-00-09&SCH=7015

Talukder, A., & Ahammed, B. (2020). Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. *Nutrition*, *78*, 110861. https://doi.org/10.1016/j.nut.2020.110861

Uganda Ministry of Health. (2003). Promoting child growth and health in Uganda: Training handbook for community child health promoters. Retrieved August 2, 2021, from http://library.health.go.ug/publications/child-health-school-health/promoting-child-growth-and-health-uganda-training-handbook

Wiens, J., & Shenoy, E. (2018). Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. *Clinical Infectious Diseases*, *66*, 149–153. https://doi.org/10.1093/cid/cix731

World Health Organization (WHO). (2012). *Combined course on growth assessment and IYCF Counselling: Trainer's guide*. Retrieved August 2, 2021, from https://apps.who.int/iris/handle/10665/77944

World Health Organization (WHO), United Nations Children's Fund (UNICEF). (2009). *WHO child growth standards and the identification of severe acute malnutrition in infants and children*. World Health Organization. Retrieved August 2, 2021, from https://apps.who.int/nutrition/publications/severemalnutrition/9789241598163/en/index.html

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nel, S., Feucht, U. D., Nel, A. L., Becker, P. J., & Wenhold, F. A. M. (2022). A novel screening tool to predict severe acute malnutrition through automated monitoring of weight-for-age growth curves. *Maternal & Child Nutrition*, 18, e13364. https://doi.org/10.1111/mcn.13364