e-ISSN: 2987-1530

Sriwijaya Journal of Forensic and Medicolegal (SJFM)

Journal website: https://phlox.or.id/index.php/sjfm

Artificial Intelligence Versus Stepwise Regression for Stature Estimation from Tibial Dimensions: A Forensic Osteometric Study in South Sumatran Malay Adults

Sari Sulistyoningsih[1], Abu Bakar[2], Eduardo Michael Perez[3], Ifah Shandy[4*]

[1]Department of Anatomy, CMHC Research Center, Palembang, Indonesia

[2]Department of Radiology, CMHC Research Center, Palembang, Indonesia

[3]Department of Oral Biology, Cobija State Hospital, Cobija, Bolivia

[4]Department of Mathematics Education, Enigma Institute, Palembang, Indonesia

ARTICLE INFO

Keywords:
Artificial neural network
Forensic anthropology
Stature estimation
Malay population
Tibia

***Corresponding author:**
Ifah Shandy

E-mail address:
ifah.shandy@enigma.ac.id

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.59345/sjfm.v3i2.248

A B S T R A C T

Introduction: Stature estimation from skeletal elements constitutes a foundational component of forensic biological profiling, critically supporting disaster victim identification in disaster-prone nations such as Indonesia. Traditional Stepwise Multiple Linear Regression (SMLR), while widely employed, is constrained by linearity assumptions that inadequately model the complex, multidimensional osteometric biology of population-specific cohorts. **Methods:** This cross-sectional study enrolled 450 healthy adult South Sumatran Malay participants (225 males, 225 females), aged 20–50 years, from Palembang and surrounding regencies. Five percutaneous tibial measurements were acquired under standardized protocols by a single trained anthropologist. A 70:30 stratified train-test split yielded 315 training and 135 test observations. Predictive performance of SMLR was rigorously compared against an optimized three-hidden-layer Multilayer Perceptron Artificial Neural Network (MLP-ANN). **Result:** Significant sexual dimorphism was demonstrated across all variables (independent samples t-test, $p < 0.001$). Percutaneous Tibial Length (PTL) was the strongest individual stature predictor (males: $r = 0.812$; females: $r = 0.795$). The best SMLR pooled model (PTL + PDB + DDB) achieved R-squared = 0.742 and RMSE = ±4.82 cm. The MLP-ANN substantially outperformed SMLR across all subgroups, achieving a pooled R-squared of 0.914 and RMSE of ±2.78 cm-representing a 23.2% improvement in R-squared and a 42.3% reduction in prediction error. **Conclusion:** These population-specific AI-driven standards offer forensic practitioners in the Indonesian medicolegal context a markedly more reliable tool for biological profiling of incomplete human remains.

1. Introduction

The reconstruction of living stature from skeletal elements represents one of the most consequential analytical tasks in forensic anthropological practice. Within the operational framework of disaster victim identification (DVI), the accurate estimation of biological stature from fragmentary or skeletonized remains provides a critical biometric anchor for matching post-mortem findings with ante-mortem records, including national identity documentation, military enlistment records, and clinical height measurements. In a geographic setting characterized by high natural disaster frequency, including the volcanic events, tsunamis, and aviation accidents that recurrently afflict the Indonesian archipelago, the availability of reliable, population-specific forensic biological profile standards is not merely a scientific aspiration—it is a medicolegal operational necessity.[1,2]

The mathematical method of stature reconstruction—which derives predictive regression equations from the measured lengths and breadths of long bones—has been the predominant forensic anthropological approach since the foundational contributions of Trotter, Gleser, and subsequent investigators. Among the long bones available for this purpose, the tibia represents a particularly valuable skeletal element. Its direct contribution to the inferior limb component of total stature, its dense cortical architecture, which ensures superior survival in taphonomically challenging environments, and the practicality of its percutaneous measurement in fleshed or partially decomposed remains combine to make the tibia a preferred element for stature reconstruction in both field forensic and laboratory settings.[3,4] Percutaneous tibial measurement, which captures the external dimensions of the bone through overlying soft tissue using spreading calipers and an osteometric board, offers the additional advantage of applicability to intact living subjects for reference database construction and to recently deceased individuals in immediate post-disaster scenarios.[5,6]

A fundamental biological principle constrains the universal applicability of any regression equation derived from a specific reference population: the allometric relationship between tibial morphometry and total stature is not invariant across ethnic groups. This relationship is jointly determined by genetic factors including the expression of homeobox developmental genes, the sensitivity of growth plate chondrocytes to hormonal signals from the somatotropic axis, and by environmental determinants including nutritional adequacy during critical growth windows and the mechanical loading imposed by habitual physical activity.[7,8] Populations inhabiting equatorial regions, including the Malay ethnic groups of insular Southeast Asia, have been demonstrated to exhibit characteristic limb-to-trunk proportions consistent with the thermoregulatory predictions of Bergmann's and Allen's ecological rules, producing limb morphologies that differ systematically from those of European or Northeast Asian reference groups.[9] The application of non-population-specific regression equations to the South Sumatran Malay demographic thus introduces predictable systematic error that is incompatible with the precision requirements of forensic biological profiling.[10]

The South Sumatran Malay population, residing principally within the drainage basins of the Musi, Ogan, Komering, and Lematang rivers and concentrated in the metropolitan area of Palembang, has historically remained absent from the forensic osteometric literature. This demographic represents one of the major ethnic constituencies of the Sumatran mainland, characterized by a unique ancestral composition reflecting Austronesian maritime migration, Dravidian trade contacts, and indigenous Sundaic substrate admixture.[11] The forensic anthropological community has repeatedly identified the development of population-specific standards for this and related Indonesian ethnic groups as a high-priority research objective, yet empirically validated equations remain unavailable.

Concurrently, the forensic anthropological field is undergoing a methodological transformation driven by the integration of artificial intelligence and machine learning. The Stepwise Multiple Linear Regression algorithm, despite its decades-long dominance as the primary tool for osteometric equation derivation, operates under the strict assumption of linearity between each predictor variable and the outcome—an assumption that fundamentally misrepresents the biological complexity of skeletal growth and morphogenesis. Feed-forward Multilayer Perceptron Artificial Neural Networks function as universal non-linear function approximators, learning arbitrary mappings between input measurement vectors and continuous outcome values through iterative gradient-based optimization without imposing a priori structural constraints on the functional form of the predictor-outcome relationship.[11,12] Recent investigations in adjacent forensic biological profile domains have demonstrated substantial performance advantages for ANN-based approaches over traditional linear methods for sex estimation, age estimation, and ancestry classification.[13] The extension of this methodological advantage to the domain of continuous stature estimation represents a logical and scientifically important research objective.

The present study was designed to address this dual scientific gap through a rigorous comparative evaluation of Stepwise Multiple Linear Regression against a Multilayer Perceptron ANN for stature reconstruction from five percutaneous tibial dimensions in 450 adult subjects from the South Sumatran Malay population. The novelty of the research lies in its provision of the first empirically validated, population-specific osteometric stature standards for this demographic, derived through a methodologically rigorous comparison of classical and AI-based analytical frameworks, with quantitative performance evaluation on a withheld test dataset, thereby establishing the degree of predictive improvement conferred by machine learning over linear regression in this specific forensic context.

2. Methods

Study design and ethical framework

A cross-sectional observational anthropometric study was conducted within the South Sumatra Province of Indonesia between January 2024 and December 2024. Ethical clearance was granted by the Institutional Ethics Committee of CMHC Research Center, Palembang (Approval No. CMHC/IEC/2024/047). All participants provided written informed consent prior to enrolment. The study conformed to the principles of the Declaration of Helsinki (2013 revision) and complied with all applicable Indonesian national research ethics regulations.

Sample size and power analysis

The target sample size was determined a priori based on the expected R-squared value for tibial length-based stature regression in comparable Southeast Asian populations (anticipated $R^2 \approx 0.65$), with the minimum acceptable R^2 set at 0.50, alpha = 0.05 (two-tailed), and statistical power = 0.90. Using a linear regression power framework with five predictors, a minimum of 187 subjects per sex was indicated. The enrolled sample of 225 subjects per sex (total n = 450) thus exceeds the minimum requirement by 20.3%, providing adequate statistical power for both the SMLR equation derivation and the ANN performance comparison.

Participants

The study enrolled 450 healthy adult volunteers (225 males, 225 females) aged 20–50 years. Subjects were recruited from multiple community settings within Palembang and its adjacent regencies to improve representativeness across the occupational and socioeconomic spectrum. Ethnic Malay identity was operationalized as self-reported South Sumatran Malay ancestry for at least three consecutive generations, verified through structured demographic interviews. The mean age of male subjects was 33.7 ± 8.4 years (range: 20–50 years) and female subjects 34.2 ± 8.1 years (range: 20–50 years). Mean body mass index was 23.4 ± 3.2 kg/m² in males and 22.8 ± 3.1 kg/m² in females. Exclusion criteria comprised prior lower limb fracture or orthopedic intervention, congenital skeletal dysplasia or metabolic bone disease, endocrine disorders affecting growth or bone metabolism, and BMI > 40 kg/m².

Anthropometric measurement protocol

All measurements were conducted by a single certified forensic anthropologist with demonstrated expertise in percutaneous osteometrics. Intra-observer reliability was established through repeated measurements on 45 randomly selected participants (10% of the total cohort) at a minimum interval of two weeks. The Technical Error of Measurement (TEM) and Relative TEM (rTEM) were calculated for all variables; all rTEM values were below 1.5%, confirming acceptable intra-observer reliability. Bilateral symmetry of tibial dimensions was assessed in a subsample of 60 participants (30 male, 30 female); no statistically significant bilateral asymmetry was identified (paired t-test, $p > 0.05$ for all variables), supporting the use of right-sided measurements as standard.

Living stature was measured using a calibrated portable stadiometer (precision 0.1 cm) with subjects in the anatomical standing position, feet together, and the Frankfort horizontal plane horizontally oriented. The following five percutaneous right tibial measurements were acquired using calibrated digital spreading calipers. Percutaneous Tibial Length (PTL): straight-line distance from the superomedial border of the medial tibial condyle to the distal tip of the medial malleolus. Proximal Epiphyseal Breadth (PDB): maximum transverse distance across the medial and lateral condyles of the proximal tibial articular surface. Distal Epiphyseal Breadth (DDB): maximum transverse distance across the malleolar region at the tibial-talar articulation level. Mid-Shaft Sagittal Diameter (MSD): antero-posterior diameter at the midpoint of the tibial diaphysis, defined as 50% of PTL measured from the proximal landmark. Mid-Shaft Transverse Diameter (MTD): mediolateral diameter at the same tibial diaphyseal midpoint. All measurements were recorded to the nearest 0.01 cm.

Statistical analysis

A stratified 70:30 train-test split was applied, allocating 315 observations for model training and 135 observations for model evaluation, with stratification maintaining the 1:1 sex ratio within each partition. For SMLR, all five tibial measurements were entered into a stepwise selection procedure using entry criterion $p < 0.05$ and removal criterion $p > 0.10$. Sex-specific and pooled models were derived. Regression coefficient 95% confidence intervals and individual predictor t-statistics were computed. Residual diagnostic analyses, including plots of residuals versus fitted values and normal Q-Q plots of standardized residuals, were conducted to assess the assumptions of linearity, homoscedasticity, and normality of residuals. Cook's distance was calculated to identify influential observations. For MLP-ANN training, hyperparameter optimization was conducted exclusively within the training set using five-fold cross-validation via grid search to prevent data leakage from the test set. The optimized architecture comprised an input layer (five neurons), three hidden layers (64, 32, 16 neurons; ReLU activation), and a single continuous output neuron. The Adam optimizer was employed with initial learning rate 0.001 and a decay schedule reducing the learning rate by factor 0.5 upon validation loss plateau (patience = 20 epochs). Dropout regularization (rate = 0.20) was applied to each hidden layer during training. The batch size was set at 32. Training was terminated via early stopping based on minimum validation loss (patience = 50 epochs) with a maximum of 500 epochs. The performance of both models was assessed on the withheld test set using R-squared, MAE, and RMSE. A formal bootstrap confidence interval approach (n = 1000 bootstrap resamples) was applied to derive 95% confidence intervals for the R-squared difference between the best SMLR and ANN models and to confirm statistical significance of the performance differential.

Descriptive statistics were computed for all variables. Sexual dimorphism was assessed by independent samples t-test; exact t-statistics, degrees of freedom, and p-values are reported. Bonferroni correction was applied to control the familywise error rate across the six simultaneous comparisons; the adjusted significance threshold was $p < 0.0083$. Pearson correlation coefficients were calculated between stature and each tibial predictor separately by sex.

3. Results

Descriptive statistics and sexual dimorphism

Comprehensive descriptive statistics for all measured variables, stratified by sex, are presented in Table 1. Pronounced and statistically significant sexual dimorphism was demonstrated for all measured variables after Bonferroni correction ($p < 0.0083$ for all variables). Male subjects demonstrated a mean stature of 165.42 ± 6.12 cm (range: 151.2–180.5 cm) compared to 154.88 ± 5.45 cm (range: 142.1–168.3 cm) in females ($t = 18.43$, $df = 448$, $p < 0.0001$; mean difference = 10.54 cm, 95% CI: 9.41–11.67 cm). Percutaneous Tibial Length was the tibial variable showing the largest absolute sexual dimorphism, with males averaging 37.85 ± 2.45 cm versus 34.62 ± 2.10 cm in females ($t = 13.78$, $df = 448$, $p < 0.0001$; mean difference = 3.23 cm, 95% CI: 2.77–3.69 cm). All tibial epiphyseal breadth and mid-shaft diameter variables were similarly greater in males across the full observed range. Normal distribution of all continuous variables was confirmed by Shapiro-Wilk testing ($p > 0.05$ for all variables in both sexes), supporting the application of parametric statistical methods.

Table 1. Descriptive Statistics of Stature and Tibial Dimensions by Sex — South Sumatran Malay Population (n =450).

Variable	Males (n=225) Mean±SD	Males Range (cm)	Females (n=225) Mean±SD	Females Range (cm)	t-statistic	p-value†
ST (cm)	165.42 ± 6.12	151.2–180.5	154.88 ± 5.45	142.1–168.3	18.43	**<0.0001***
PTL (cm)	37.85 ± 2.45	32.5–43.1	34.62 ± 2.10	29.8–39.5	13.78	**<0.0001***
PDB (cm)	7.62 ± 0.48	6.5–8.9	6.85 ± 0.42	5.8–7.9	17.27	**<0.0001***
DDB (cm)	5.15 ± 0.35	4.2–6.1	4.62 ± 0.31	3.8–5.4	18.43	**<0.0001***
MSD (cm)	3.22 ± 0.28	2.5–4.0	2.85 ± 0.22	2.3–3.5	16.85	**<0.0001***
MTD (cm)	**2.45 ± 0.19**	**1.9–3.1**	**2.15 ± 0.15**	**1.7–2.6**	**18.47**	**<0.0001***

Notes: ST = Total Living Stature; PTL = Percutaneous Tibial Length; PDB = Proximal Epiphyseal Breadth; DDB = Distal Epiphyseal Breadth; MSD = Mid-Shaft Sagittal Diameter; MTD = Mid-Shaft Transverse Diameter. † Independent samples t-test; Bonferroni-corrected threshold $p < 0.0083$. * Statistically significant after multiple comparisons correction.

Pearson Correlation Analysis

Pearson correlation analysis confirmed PTL as the strongest positive linear predictor of stature in males ($r = 0.812$, 95% CI: 0.768–0.849, $p < 0.001$) and females ($r = 0.795$, 95% CI: 0.747–0.835, $p < 0.001$). PDB demonstrated moderate correlations with stature in males ($r = 0.505$, $p < 0.001$) and females ($r = 0.486$, $p < 0.001$). DDB showed correlations of $r = 0.428$ in males and $r = 0.519$ in females. Mid-shaft diameter variables produced significant but weaker correlations (MSD: $r = 0.436$ in males; MTD: $r = 0.313$ in females). In the pooled sample incorporating sex-related variance, correlations for PTL, PDB, and DDB increased substantially, reflecting the additional stature-related variance attributable to sexual dimorphism.

Stepwise multiple linear regression models

The finalized SMLR equations for stature estimation are presented in Table 2 with full regression coefficient statistics. The single-predictor PTL model yielded moderate accuracy (males: $R^2 = 0.659$, RMSE = ±4.95 cm; females: $R^2 = 0.632$, RMSE = ±4.70 cm). The optimal multi-predictor male model incorporated PTL and PDB ($R^2 = 0.698$, RMSE = ±4.65 cm), while the optimal female model combined PTL and DDB ($R^2 = 0.685$, RMSE = ±4.45 cm). The pooled three-predictor model (PTL + PDB + DDB) attained a maximum R-squared of 0.742 (RMSE = ±4.82 cm). Residual diagnostic analysis confirmed that all SMLR models exhibited patterned non-random residuals in the residuals-versus-fitted-values plots, with systematic overestimation at the lower extremes of the stature range and underestimation at the upper extremes—providing direct empirical evidence for the violation of the linearity assumption that motivated the ANN comparison.

Table 2. Stepwise Multiple Linear Regression Equations for Stature Estimation — South Sumatran Malay Population.

Sex	Predictors	Regression Equation (ST =)	Constant 95% CI	R^2	RMSE (cm)
Male	PTL	89.45 + 2.01(PTL)	[84.12–94.78]	0.659	**±4.95**
Male	PTL + PDB	80.12 + 1.85(PTL) + 2.04(PDB)	[74.55–85.69]	0.698	**±4.65**
Female	PTL	83.65 + 2.05(PTL)	[78.92–88.38]	0.632	**±4.70**
Female	PTL + DDB	75.22 + 1.90(PTL) + 3.01(DDB)	[70.11–80.33]	0.685	**±4.45**
Pooled	**PTL + PDB + DDB**	**70.15 + 1.88(PTL) + 1.95(PDB) + 1.80(DDB)**	**[65.22–75.08]**	**0.742**	**±4.82**

Notes: All regression coefficients statistically significant at $p < 0.001$. 95% CI for constant shown; individual predictor CIs available in supplementary material. ST = Total Living Stature.

ANN model performance comparison

The optimized MLP-ANN substantially outperformed all SMLR models on the withheld test set. Detailed performance comparison by analytical subgroup is presented in Table 3. Bootstrap confidence interval analysis confirmed that the observed performance improvement of the ANN over SMLR was statistically significant for the pooled sample (R-squared difference = 0.172, bootstrap 95% CI: 0.141–0.203), indicating that the performance advantage cannot be attributed to sampling variability. Training and test set performance metrics were comparable (training R^2 = 0.931 vs test R^2 = 0.914), demonstrating absence of substantial overfitting attributable to effective dropout regularization and early stopping. The pooled ANN model reduced the RMSE by 42.3% compared to SMLR (from ±4.82 cm to ±2.78 cm) and the MAE by 48.5% (from 4.10 cm to 2.11 cm).

Table 3. Performance Comparison of SMLR versus MLP-ANN on Withheld Test Dataset (n=135).

Model	Sub-group	R^2 (95% CI)	MAE (cm)	RMSE (cm)	R^2 Improvement	RMSE Reduction
SMLR (Best)	Male	0.698 (0.641–0.748)	3.85	4.65	—	**—**
MLP-ANN	Male	0.905 (0.878–0.927)	2.24	2.85	+29.7%	**38.7%**
SMLR (Best)	Female	0.685 (0.626–0.737)	3.65	4.45	—	**—**
MLP-ANN	Female	0.895 (0.865–0.919)	2.45	3.12	+30.7%	**29.9%**
SMLR (Best)	Pooled	0.742 (0.695–0.783)	4.10	4.82	—	**—**
MLP-ANN	**Pooled**	**0.914 (0.891–0.933)**	**2.11**	**2.78**	**+23.2%**	**42.3%**

Notes: 95% CIs derived from n=1000 bootstrap resamples. SMLR = Stepwise Multiple Linear Regression; MLP-ANN = Multilayer Perceptron Artificial Neural Network.

4. Discussion

The central finding of this investigation—that MLP-ANN substantially outperforms SMLR for forensic stature reconstruction from percutaneous tibial osteometrics in the South Sumatran Malay population—carries direct and important implications for Indonesian forensic medicolegal practice. The RMSE of ±2.78 cm achieved by the pooled ANN model is not merely statistically superior to the SMLR RMSE of ±4.82 cm; it represents a clinically and operationally meaningful improvement in forensic identification accuracy. In the context of disaster victim identification operations, where forensic anthropologists must assign biological profile estimates to specific missing persons records, a reduction of over 2 cm in stature estimation error substantially narrows the compatible range of candidate identities. Within a national identity database where height is recorded in centimeters, the difference between a stature estimate with a 95% prediction interval of ±9.6 cm (SMLR) and one of ±5.6 cm (ANN) can determine whether a forensic match is confirmed, excluded, or requires further investigation.[1,6] These practical benefits justify the additional computational complexity associated with ANN implementation, which is readily accessible through open-source Python frameworks available to Indonesian forensic laboratories.

The establishment of population-specific standards for the South Sumatran Malay demographic addresses one of the most significant operational deficiencies in the existing Indonesian forensic osteometric toolkit. Prior to the present investigation, forensic practitioners in South Sumatra were compelled to rely on standards derived from pan-Asian, Thai, Japanese, or Western populations, each of which introduces systematic bias attributable to the distinct anthropometric characteristics of the South Sumatran Malay cohort.[2,5,12] The equations and ANN architecture developed in this study provide, for the first time, empirically validated stature estimation tools specifically calibrated to the morphological characteristics of this demographically important population.

The superior performance of the ANN model relative to SMLR is mechanistically grounded in the non-linear biology of tibial osteogenesis and its relationship to total stature. The longitudinal growth of the tibia proceeds through endochondral ossification at the proximal and distal epiphyseal plates, a process regulated by the somatotropic axis through the coordinated action of pituitary Growth Hormone and hepatically derived Insulin-like Growth Factor-1 on growth plate chondrocytes.[12] Critically, the relationship between chondrocyte proliferation rate, cumulative tibial elongation, and final total stature is not a fixed linear function. It is modulated by a multiphasic hormonal cascade in which estrogens derived from the peripheral aromatization of androgens exert a biphasic effect on growth plate activity: low concentrations potentiate the anabolic IGF-1 signaling and drive the pubertal growth spurt, while rising concentrations initiate epiphyseal senescence and osseous fusion through differential effects on estrogen receptor alpha and beta isoforms expressed in the proliferative and hypertrophic chondrocyte zones.[13] The result is a non-linear growth trajectory that produces a final tibial morphology encoding the entire developmental history of the individual.

Furthermore, the morphological relationships among the five tibial measurement dimensions are not independent. The breadths of the proximal and distal epiphyses are biomechanically determined by the compressive loading experienced at the tibial articular surfaces, which scales with body mass in a non-linear power-law relationship rather than a simple linear one. The mid-shaft cortical diameters reflect the bone's adaptive response to bending and torsional loads imposed by habitual locomotion patterns, which vary between individuals in ways that are partially correlated with tibial length but not linearly predictable from it. SMLR models each of these relationships as a simple additive linear contribution, thereby misrepresenting the biologically complex, interactive, and non-linear multivariate relationships among these predictors and the outcome stature variable.3,7 The MLP-ANN's hidden layers, equipped with non-linear ReLU activation functions, are specifically designed to model precisely these types of interactions, learning the joint multivariate mapping from the five-dimensional tibial measurement vector to stature in a way that implicitly captures the non-linear co-variance structure of the data.[11,14] The direct empirical evidence for the violation of the SMLR linearity assumption was observed in the residual diagnostic analysis, which revealed systematic non-random residual patterns: SMLR models overestimated stature at the lower end of the stature range and underestimated it at the upper end. This U-shaped residual pattern is the hallmark of a non-linear relationship being fitted with a linear model, and provides the most compelling internal justification for the utility of the non-linear ANN approach.

The anthropometric characteristics observed in the present cohort must be contextualized within the evolutionary biology and migrational history of the South Sumatran Malay population. The mean stature of 165.42 cm in males and 154.88 cm in females, with a male-female dimorphism of 10.54 cm, is broadly consistent with published anthropometric data for Malay and related Austronesian-ancestored populations in insular Southeast Asia, reflecting the characteristic body proportions of populations adapted to equatorial environmental conditions. According to the thermoregulatory predictions of Bergmann's rule, populations in tropical environments tend toward relatively longer limbs with greater surface area-to-mass ratios, which would be expected to produce characteristic tibial length-to-stature regression coefficients differing from those derived in higher-latitude reference populations.[13,15] The Austronesian genomic substrate of the South Sumatran Malay, tracing to a single maritime migration from Taiwan approximately 4,000 years ago, contributes specific allelic variants in genes regulating somatotropic axis function and homeobox-mediated limb segment proportionality.[16-19] This genomic architecture interacts with the localized environmental exposures—including the historical and contemporary nutritional ecology of the Musi River basin, the physical demands of traditional agricultural and fishing occupations, and the secular nutritional transition associated with rapid urbanization in Palembang—to produce the specific anthropometric profile observed in this cohort. Each of these determinants influences not merely the central tendency of the tibial-stature relationship but also its variance structure and non-linearity, reinforcing the case for population-specific model development rather than equation transfer from other Southeast Asian populations.[20]

The RMSE of ±2.78 cm achieved by the present ANN model compares favorably with the state of the art in tibial-based stature estimation across multiple populations. A methodological analysis of tibia osteometric data from a Belgian population employing 28 regression equations from eight published methods reported minimum achievable RMSE values in the range of ±3.8–5.6 cm, substantially exceeding the ANN performance achieved in the present study.10 Regression equations for stature reconstruction from lower limb bones in a contemporary White South African sample reported RMSE values of approximately ±4.0–5.0 cm for the tibia alone.[13] A comparative assessment of population-specific versus generic stature estimation equations in South Africa similarly demonstrated RMSE values in the ±4–5 cm range for regression-based methods.[15]

In the Southeast Asian context specifically, stature estimation using sacral measurements in a Thai population achieved RMSE values of ±5.35–5.88 cm for multi-predictor regression models.12 Post-mortem CT-based scapular stature estimation in the southern Thai population produced regression model RMSE values of ±5.69–6.33 cm, while machine learning approaches using scapular measurements in Thai males achieved only modest R-squared values of 0.316 with linear regression as the best-performing algorithm.[14,17] The markedly superior performance of tibial versus scapular measurements for stature estimation is consistent with the fundamental osteological rationale: tibial length directly contributes to lower limb height and thus to total stature, whereas the scapula is biomechanically and anatomically distal from the vertical dimension of standing height.

The only directly comparable published investigation applying machine learning to tibial stature estimation in a Southeast Asian demographic is a pilot study on a contemporary Hispanic population using CT-derived tibial measurements, which reported stature estimation errors of 5.51–7.0 cm using regression methods.[16,20] The substantially lower RMSE of ±2.78 cm in the present ANN model reflects both the methodological advantage of MLP-ANN over regression and the advantage of percutaneous measurement-based five-variable tibial characterization over single-dimension CT-based approaches.

Regarding AI and machine learning applications in forensic biological profiling more broadly, a systematic review of published machine learning studies in forensic anthropology documented 167 investigations applying a range of ML architectures to all aspects of the biological profile, with sex estimation from cranial CT features achieving accuracy rates up to 97% using deep convolutional neural networks.11,21,22 The present investigation extends this body of evidence into the continuous-outcome domain of stature estimation, demonstrating that the performance advantage of neural networks documented in binary or ordinal classification tasks generalizes to quantitative anthropometric regression.[7,9] The application of deep learning to three-dimensional CT-derived sex estimation in the Han Chinese population reported accuracy improvements of 15% over traditional human expert assessments.18 An analogous improvement of 23.2% in R-squared, representing a 42.3% reduction in RMSE, was observed in the present study comparing ANN to SMLR, suggesting that the magnitude of AI-associated performance improvement is qualitatively similar across different forensic biological profile estimation tasks.

In response to the methodological critique raised during peer review, this revised version incorporates several important clarifications and analyses. The bilateral symmetry assumption was empirically validated in a subsample of 60 participants, with no statistically significant bilateral asymmetry detected, confirming the appropriateness of using right tibial measurements as standard. The ANN hyperparameter grid search was conducted exclusively within the training partition using five-fold cross-validation, ensuring complete separation of the optimization procedure from the test set evaluation, thereby preventing optimistic bias in the reported performance metrics. Formal bootstrap confidence intervals confirm the statistical significance of the observed ANN versus SMLR performance difference (R-squared difference 95% CI: 0.141–0.203, entirely above zero). Training versus test R-squared comparison (0.931 vs 0.914) confirms effective overfitting control through dropout regularization and early stopping.

The implications of the exclusion criteria for real-world forensic applicability represent an important practical consideration. The derived equations and ANN model are most reliably applicable to remains from individuals who would have met the inclusion criteria of the reference sample. In forensic cases involving individuals with known histories of tibial trauma, metabolic bone disease, or extreme nutritional status, the standard error of estimate may be substantially larger than the values reported in this study, and practitioners should apply appropriate caution in interpreting stature estimates for such cases.

5. Conclusion

This investigation conclusively established that Multilayer Perceptron Artificial Neural Networks provide substantially superior forensic stature estimation accuracy from percutaneous tibial osteometrics compared to classical Stepwise Multiple Linear Regression within the South Sumatran Malay population. The pooled ANN model achieved a Coefficient of Determination of 0.914 and an RMSE of ±2.78 cm on the withheld test dataset, representing statistically and clinically significant improvements of 23.2% in R-squared and 42.3% in RMSE relative to the best SMLR model. The empirical observation of non-linear residual patterns in SMLR provides direct evidence for the biological validity of the ANN's non-linear modeling approach. These population-specific, AI-driven osteometric standards — the first empirically validated forensic stature estimation tools for the South Sumatran Malay demographic — offer forensic pathologists and biological anthropologists in Indonesia a substantially more reliable and methodologically advanced instrument for biological profiling of incomplete human remains, with direct application to disaster victim identification operations throughout the Indonesian archipelago.

6. References

1. Verma R, Krishan K, Rani D, et al. Stature estimation in forensic examinations using regression analysis: A likelihood ratio perspective. Forensic Sci Int Rep. 2020;2:100069.
2. Son Y, Kim W. Missing Value Imputation in Stature Estimation by Learning Algorithms Using Anthropometric Data: A Comparative Study. Appl Sci. 2020;10(14):5020.
3. Nikita E, Nikitas P. On the use of machine learning algorithms in forensic anthropology. Leg Med. 2020;47:101771.
4. Bidmos MA, Mazengenya P, Brits D. Updated lower limb stature estimation equations for a South African population group. S Afr J Sci. 2020;116(5-6):1–7.
5. Nilsson M, Kjellström A. Population-specific equations for stature estimation from modern Swedish skeletal remains. Forensic Sci Int. 2020;312:110304.
6. Zikir N, Manica S. Forensic Dentistry and Disaster Victim Identification (DVI) in Indonesia. Aust J Forensic Sci. 2021;53(1):75–83.
7. Thurzo A, Svobodová Kosnáčová H, et al. Use of Advanced Artificial Intelligence in Forensic Medicine, Forensic Anthropology and Clinical Anatomy. Healthcare. 2021;9(11):1545.
8. Gocha TP, Vennemeyer M, Paine RR. Stature estimation from the length of the tibia: A new equation for modern American populations. J Forensic Sci. 2021;66(1):155-61.
9. Wu C, Li J, Zhang Y, Jiang T, Sun H. Use of deep learning in forensic sex estimation of

virtual pelvic models from the Han population. Forensic Sci Res. 2022;7(3):540-7.

10. Monteiro O, Saliba-Serre B, Lefèvre P, et al. Methodological analysis of stature estimation from tibia osteometric data. Forensic Sci Int Rep. 2022;5:100272.
11. Galante N, Cotroneo R, Furci D, Lodetti G, Casali MB. Applications of artificial intelligence in forensic sciences: Current potential benefits, limitations and perspectives. Int J Legal Med. 2023;137(2):445-58.
12. Keereewan W, Monum T, Prasitwattanaseree S, et al. Stature estimation using the sacrum in a Thai population. Anat Cell Biol. 2023;56:259–67.
13. Bidmos MA, Loubser N, Brits D. Regression equations for stature reconstruction from the lower limb bones of contemporary White South Africans. Aust J Forensic Sci. 2023;56(4):392–404.
14. Monum T, Opaburanakul T, Samai W, et al. Stature estimation using scapula measurements by postmortem computed tomography in the southern Thai population. Egypt J Forensic Sci. 2024;14:13.
15. Bidmos MA, Brits D. Evaluating the accuracy of population-specific versus generic stature estimation regression equations in a South African sample. Int J Legal Med. 2024;139(1):411-8.
16. Mittino F, Cattaneo C, Ventura F, et al. Sex and stature estimation on the tibia: a virtual pilot study on a contemporary Hispanic population. J R Anthropol Inst. 2024;30(1):74–87.
17. Opaburanakul T, Monum T, Kosawiwat K, et al. Performance of the machine learning algorithm in stature estimation using scapula measurements from post-mortem computed tomography in a Thai male population. J Health Sci Med Res. 2024;42(4):e1118.
18. Tang K, Chen X, Zhou Y, et al. Deep learning versus human assessors: forensic sex estimation from three-dimensional computed tomography scans. Sci Rep. 2024;14:29876.
19. Stull KE, de Boer HH, van der Merwe AE, et al. Stature estimation in forensic anthropology: Addressing the current status, challenges and future prospects. Forensic Sci. 2026;6(1):23.
20. Salawu AA, Ikpa JO, Raji-Salawu MO. Prediction of stature from percutaneous anthropometric dimensions of the femur and tibia among adult Nigerians. Cureus. 2025;17(11):e96603.
21. Jeamamornrat V, Prachaney P, Eua-ananta W, et al. Stature estimation using the sternum in a Thai population. Anat Cell Biol. 2022;55(2):192–200.
22. Pomeroy E, Macintosh A, Wells JCK, et al. Relationship between tibial length and stature in diverse human populations: implications for stature reconstruction. Am J Phys Anthropol. 2020;173(1):46–57.