

JAST (Journal of Animal Science and Technology) TITLE PAGE

Upload this completed form to website with submission

ARTICLE INFORMATION	Fill in information in each box below
Article Type	Research article
Article Title (within 20 words without abbreviations)	Standardization of backfat thickness measurements across A-mode and B-mode ultrasound systems in pigs
Running Title (within 10 words)	Standardizing pig backfat measurements across A- and B-mode ultrasound systems
Author	Dong-Hun Noh ¹ , Mun-Hye Kang ² , Ji-Won Kim ³ , Sung Woon Choi ⁴ , Jong-Hyun Jung ⁵ , Sang-Min Lee ³ , Chang-Gwon Dang ³ , Byoung-Ho Park ³ , Sang-Hyon Oh ^{1,6}
Affiliation	¹ Division of Animal Science, Gyeongsang National University, Jinju 52828, Korea ² Division of Aerospace and Software Engineering, Gyeongsang National University, Jinju 52828, Korea ³ National Institute of Animal Science, RDA, Cheonan 21000, Korea ⁴ Nonghyup Agribusiness Group Pig Breeding Center, Gochang 56430, Korea ⁵ Jung P&C Institute, Yongin 16950, Korea ⁶ Regional Animal Industry Center, Gyeongsang National University, Jinju 52828, Korea
ORCID (for more information, please visit https://orcid.org)	Dong-Hun Noh https://orcid.org/0009-0009-6530-8464 Mun-Hye Kang https://orcid.org/0009-0003-3252-9339 Ji-Won Kim https://orcid.org/0000-0002-8417-1477 Sung Woon Choi https://orcid.org/0009-0008-6321-4782 Jong-Hyun Jung https://orcid.org/0000-0003-3667-7710 Sang-Min Lee https://orcid.org/0000-0001-8549-979X Chang-Gwon Dang https://orcid.org/0000-0003-1026-0167 Byoung-Ho Park https://orcid.org/0000-0001-7078-9463 Sang-Hyon Oh https://orcid.org/0000-0002-9696-9638
Competing interests	No potential conflict of interest relevant to this article was reported.
Funding sources State funding sources (grants, funding sources, equipment, and supplies). Include name and number of grant if available.	This project was carried out as a commissioned research project of the National Institute of Animal Science (Project No. 202403940001).
Acknowledgements	Not applicable.
Availability of data and material	Upon reasonable request, the datasets of this study can be available from the corresponding author.
Authors' contributions Please specify the authors' role using this form.	Conceptualization: SHO, BHP Data curation: DHN, JWK, SWC, JHJ, SHO Formal analysis: DHN, MHK, JWK, SHO Methodology: DHN, MHK, SWC, JHJ, SML, CGD, BHP, SHO Software: DHN, MHK, SHO Validation: DHN, MHK, SWC, SML, CGD, BHP, SHO Investigation: DHN, MHK, JWK, SHO Writing - original draft: DHN, MHK, SHO Writing - review & editing: DHN, MHK, JWK, SWC, JHJ, SML, CGD, BHP, SHO
Ethics approval and consent to participate	This article does not require IRB/IACUC approval because there are no human and animal participants.

CORRESPONDING AUTHOR CONTACT INFORMATION

For the corresponding author (responsible for correspondence, proofreading, and reprints)	Fill in information in each box below
First name, middle initial, last name	Prof. Sang-Hyon OH
Email address - this is where your proofs will be sent	shoh@gnu.ac.kr
Secondary Email address	
Address	Division of Animal Science, Gyeongsang National University, Jinju 52828, Korea
Cell phone number	
Office phone number	+82-55-772-1955
Fax number	+82-55-772-3689

Standardization of backfat thickness measurements across A-mode and B-mode ultrasound systems in pigs

Abstract

Backfat thickness is a key carcass trait used for economic valuation and genetic selection in pigs. However, field datasets are often generated using heterogeneous ultrasound devices, making direct comparison and integration difficult. This study evaluated the practical compatibility between an A-mode ultrasound device (PIGLOG 105) and a B-mode ultrasound device (ExaGo) for backfat thickness measurement and developed calibration/conversion models to standardize measurements to carcass reference scales. A total of 69 Yorkshire pigs from a domestic GGP farm were evaluated during July–August 2024. A-mode ultrasound measurements were collected 1–2 days prior to harvest at the 10–11th thoracic region. B-mode images were acquired at the identical location in two scan orientations: vertical (cross-sectional) and parallel (longitudinal). Images were stored in JPEG format and quantified using both Python-based analysis and ImageJ. Reference measurements included manual backfat thickness on hot carcasses (immediately post-harvest) and cold carcasses (after 24h chilling), VCS2000 outputs, and a direct measurement on the loin cut surface at the same anatomical level. Associations were assessed using Pearson correlation and simple linear regression with a zero intercept, and systematic bias was tested using paired *t*-tests. B-mode measurements showed closer agreement with reference values under vertical than parallel scanning. Slopes did not differ from unity for comparisons against the loin cut surface, which was measured at the same anatomical site, in three of the four B-mode combinations, whereas the smallest prediction error was observed for Python-vertical versus cold carcass. Despite high linearity, paired comparisons indicated significant systematic bias for VCS2000 and carcass references ($p < 0.001$), implying that raw B-mode values require calibration before pooling with reference-

scale data. A-mode values were stably convertible to B-mode across four combinations (Python/ImageJ × vertical/parallel) using a single coefficient ($\beta = 1.107\text{--}1.268$; RMSE = 0.233–0.362). Direct A-mode prediction of reference measurements yielded slopes of 1.199–1.879, all significantly greater than unity ($p < 0.001$), and raw A-mode differed significantly from all references (mean differences 0.276–1.245cm; $p < 0.001$). After applying regression-based calibration, mean differences were no longer significant ($p = 0.142\text{--}0.528$), leave-one-out cross-validation showed prediction errors only 0.6–1.1% larger than fitted errors, and concordance improved substantially (CCC 0.092–0.521 to 0.483–0.745). In conclusion, A-mode and B-mode are not directly interchangeable as raw data, but become compatible after applying the proposed conversion and calibration models, enabling practical standardization of field ultrasound measurements to carcass reference scales.

Keywords: backfat thickness, A-mode, B-mode, ultrasound, compatibility, pig

Introduction

Backfat thickness in pigs is a representative indicator of carcass fat deposition and is extensively utilized as a key production trait in carcass value evaluation and breeding selection programs [1, 2]. It is commonly considered in conjunction with growth performance, body composition, and fat-related meat quality traits, and its associations with productivity and meat quality may vary depending on breeding objectives and breed-specific characteristics [2, 3]. Accurate *in vivo* measurement of backfat thickness enables early selection among individuals without reliance on post-slaughter data, thereby improving the efficiency of selection decisions. This highlights its practical importance in pig breeding programs [4, 5].

However, measurement variability in field conditions may arise even within the same individual due to differences in equipment, measurement sites, scanning orientation, operator proficiency, and environmental factors [6]. These sources of variation can compromise the reliability of individual phenotypic evaluation and consequently reduce the consistency and accuracy of selection decisions. Therefore, systematic validation of the relationship between *in vivo* ultrasound measurements and actual carcass measurements is essential, along with the establishment of standardized measurement protocols to ensure reproducible and reliable data across diverse conditions [4, 6].

Among non-invasive methods for measuring backfat thickness, ultrasound devices are the most widely used and, in field practice, A-mode and B-mode systems are often used either individually or in combination [7, 8]. A-mode ultrasound, which is based on single-signal output, enables rapid numerical estimation and thus offers high practicality under field conditions. Devices such as the PIGLOG 105 have been extensively utilized in practical settings for many years [9, 10]. In contrast, B-mode ultrasound provides two-dimensional images that allow visual differentiation between fat and muscle layers, enabling more precise anatomical assessment. It has also been reported to be useful not only for measuring backfat

thickness but also for predicting various carcass traits, including loin eye area [7, 11, 12]. Furthermore, the predictive performance of B-mode measurements may vary depending on scanning orientation (transverse vs. longitudinal) and image analysis methods, highlighting the importance of well-defined measurement and analysis protocols [13, 14]. A key challenge arises from the concurrent use of A-mode and B-mode devices in commercial farms and breeding programs, where heterogeneous measurement systems may generate “numerically incompatible” data for the same trait within a single database. If device-specific biases are not properly adjusted, this may lead to inconsistencies in data integration across farms, genetic evaluation, and selection decision-making [6, 7]. This issue is particularly critical in national breeding programs, where data standardization is essential, and thus the validation of inter-device compatibility and the establishment of conversion systems should be prioritized [15]. In this context, bridging the high-precision, image-based measurements provided by B-mode with the field applicability and numerical outputs of A-mode is essential. Such integration would enable the continued utilization of existing datasets while simultaneously ensuring long-term consistency and reliability in measurement systems.

Accordingly, this study was conducted to evaluate the compatibility of backfat thickness measurements obtained from an A-mode ultrasound device (PIGLOG 105) and a B-mode ultrasound device (ExaGo), which are widely used in field conditions in Korea, and to develop a conversion model for data standardization. Backfat thickness was measured in the same individuals using both A-mode and B-mode ultrasound (with scanning orientation considered), and the measurements were compared with post-slaughter carcass values. The relationships among these measurements were assessed using correlation and regression analyses. Specifically, this study was designed to (i) first examine the relationship between B-mode measurements and carcass measurements (B→carcass), (ii) establish a conversion step to predict B-mode measurements from A-mode measurements (A→B), and (iii) subsequently

validate the predictive capability of A-mode measurements for carcass values (A→carcass). Through this stepwise approach, the study aimed to empirically demonstrate whether heterogeneous field data can be integrated into a unified standard scale. Ultimately, this study seeks to provide a practical basis for the consistent accumulation and utilization of comparable backfat thickness data in environments where multiple ultrasound systems coexist, thereby improving the efficiency of breeding selection and genetic improvement programs.

Materials and Methods

Animals and Experimental Design

This study was conducted on 69 Yorkshire pigs raised at a domestic GGP farm in Korea between July and August 2024. Data collection was carried out over three sessions (July 18, August 9, and August 30, 2024). *In vivo* ultrasound measurements and body weight recordings were performed 1–2 days prior to harvest. Subsequently, the same individuals were harvested to obtain carcass measurements, followed by statistical analysis. The overall workflow of the study—comprising tests for significant differences among measurement modes and carcass values, development of correction and prediction equations, and evaluation of compatibility—was conducted according to the procedure illustrated in Figure 1.

Ultrasound Measurements and Image Acquisition

Backfat thickness was measured using an A-mode ultrasound device (PIGLOG 105, Carometec, Denmark) and a B-mode ultrasound device (ExaGo, ECM, France). Prior to ultrasound measurements, hair was removed from the region between the 10th and 13th ribs, and vegetable oil was applied to ensure proper probe contact. For A-mode measurements, backfat thickness was recorded numerically at the 10th–11th rib region, and three repeated measurements were obtained per individual. A trained operator identified the 10th–11th thoracic vertebrae region by manual palpation and rib counting to ensure consistent anatomical

localization across individuals. For B-mode measurements, images were acquired at the same anatomical location (10th–11th ribs) using two scanning orientations: (1) a longitudinal view (parallel to the spine; parallel) and (2) a cross-sectional view (perpendicular to the spine; vertical). For each orientation, three images per individual were collected and stored in JPEG format. In addition, scans were performed across the measurement region (10th–13th ribs) at approximately 7cm from the dorsal midline (at least three images per individual), and probe angle was carefully maintained to minimize measurement error.

Vertical scanning was introduced from the second data collection session onward; consequently, vertical images were available for 35 of the 69 animals, whereas parallel images were available for 59. In addition, loin cut surfaces could not be obtained for 15 animals owing to carcass processing constraints at the abattoir, and cold carcass measurements were unavailable for three animals. All analyses were therefore based on pairwise complete cases, and the corresponding sample sizes are indicated in each table.

Digital Image Analysis

B-mode ultrasound images were analyzed using ImageJ and Python. Pixel measurements were converted to millimeters based on the reference scale embedded in each image, and backfat thickness was subsequently quantified. All ultrasound measurements and image analyses were performed by a single trained operator following a standardized protocol to minimize operator-related variability.

In the Python-based analysis, B-mode ultrasound images were processed using the OpenCV library. The workflow consisted of image loading and preprocessing, automatic contour detection to identify the upper and lower boundaries of the backfat layer, calculation of the pixel distance between detected contours, and conversion of pixel distances into millimeters using the reference scale embedded in each image.

Carcass Measurements (Reference Values)

All animals were harvested following ultrasound measurements. Carcass backfat thickness was collected using two approaches: manual measurement and automated system-based measurement. For manual measurements, backfat thickness was recorded at three standardized anatomical locations (shoulder, between the 10th and 11th thoracic vertebrae, and hip region) on the right side of the carcass, and the mean value was used for analysis. Measurements were first taken immediately post-harvest under hot carcass conditions, and subsequently repeated at the same location after 24h of chilling under cold carcass conditions.

In addition, backfat thickness data for the same individuals were obtained using an automated carcass grading system, VCS2000 (Video Carcass Grading System, E+V Technology, Germany), installed at the abattoir to evaluate practical compatibility under industrial conditions. This allowed assessment of the extent to which *in vivo* ultrasound measurements correspond not only to manual carcass measurements but also to automated measurements used in commercial grading systems. Furthermore, the loin (longissimus dorsi) region for each individual was excised at the same anatomical level (between the 10th and 11th thoracic vertebrae), and the cross-sectional surface was exposed to allow direct measurement of backfat thickness.

Statistical Analysis and Development of Conversion (Calibration) Models

Backfat thickness values extracted from B-mode images were processed using Python 3.8, and statistical analyses were performed using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). Relationships among measurement modes and carcass values were evaluated using Pearson correlation coefficients (r) and coefficients of determination (R^2). Pearson correlation coefficients and coefficients of determination are reported as descriptive summaries of the

linear association and were not used as criteria of measurement agreement; agreement was assessed by the test of the slope against unity, Lin's concordance correlation coefficient, and Bland–Altman analysis. Simple linear regression analyses were conducted to convert A-mode measurements to B-mode or carcass values.

A regression model without an intercept (i.e., forced through the origin) was applied to ensure inter-device compatibility and to standardize the data to the level of carcass measurements. This approach reflects physical plausibility while simplifying the model, allowing data conversion using a single calibration coefficient. Root mean square error (RMSE) was also calculated to compare predictive accuracy among measurement methods. Following the analytical framework illustrated in Figure 1, (1) differences between B-mode measurements and carcass values were first tested for statistical significance, and B→carcass calibration and prediction equations were developed if necessary; (2) an A→B conversion model was then established; and (3) an A→carcass prediction model was subsequently derived. Because the aim of these analyses was to determine whether measurements obtained from different systems are equivalent, rather than to estimate the relationship between two variables, models were fitted through the origin and the slope was the parameter of interest, with the null hypothesis $H_0: \beta = 1$ corresponding to the identity line ($y = x$). In the B-mode step, B-mode values were fitted as the dependent variable against the reference measurements (Table 2, Fig. 2), so that the slope expresses the deviation of B-mode readings from the reference scale; in the A→B and A→reference steps, A-mode values were the independent variable. Because the hypothesis tested is equivalence to the identity line, the direction of fitting does not affect the null value of unity, although the magnitude of an estimated slope is direction-dependent and the slopes in Table 2 are therefore not the reciprocals of those in Table 5. The robustness of the calibration models was assessed by leave-one-out cross-validation, and the slope test was repeated using leave-one-out calibration coefficients to avoid circularity arising from coefficient estimation

on the same dataset. Homoscedasticity was examined using White's specification test. The predictive capability of A-mode measurements for carcass values and the compatibility between A- and B-mode systems were evaluated based on this stepwise approach. Differences between modes were assessed using paired *t*-tests.

Results

Descriptive Statistics of General Performance and Backfat Thickness Measurements

Descriptive statistics for *in vivo* ultrasound measurements and carcass measurements of Yorkshire pigs included in this study are presented in Table 1. The mean carcass weight and live body weight were 90.02 ± 4.047 kg and 120.6 ± 4.723 kg, respectively. For backfat thickness measured *in vivo*, the A-mode measurement showed a mean value of 1.297 cm, whereas B-mode measurements varied depending on the analysis method and scanning orientation.

When an automated algorithm (Python) was applied, the vertical (perpendicular to the spine) and parallel (parallel to the spine) measurements were 1.611 cm and 1.576 cm, respectively. In contrast, manual measurements using image analysis software (ImageJ) yielded slightly lower values of 1.512 cm and 1.475 cm for the vertical and parallel orientations, respectively. Overall, B-mode measurements tended to be higher than A-mode measurements regardless of the analysis method, although ImageJ-based values were consistently slightly lower than those obtained using the Python-based automated approach.

Post-slaughter measurements showed that backfat thickness was highest under hot carcass conditions (mean = 2.542 cm), whereas it decreased to 2.191 cm after chilling, with measurements taken 24 h post-slaughter while the carcasses were suspended under cold carcass conditions, reflecting cooling-related shrinkage. The automated carcass grading system (VCS2000) yielded a mean backfat thickness of 1.896 cm, whereas a value of 1.531 cm was

obtained by direct measurement from the loin cross-section collected at the 10th–11th rib region of the carcass.

In summary, differences in absolute backfat thickness values were observed depending on the measurement method and timing (*in vivo* vs. carcass). These discrepancies are likely attributable to differences in measurement principles among devices, analytical algorithms, and physical changes associated with carcass chilling.

Relationship between B-mode Ultrasound Measurements and Carcass Measurements

Backfat thickness values extracted from B-mode ultrasound images (vertical and parallel scans) were evaluated against carcass measurements (hot and cold carcass) and reference measurements (VCS2000 and loin cross-section) using simple linear regression and correlation analyses (Fig. 2). Overall, the vertical orientation showed lower prediction error (RMSE) than the parallel orientation, with the Python-based vertical analysis demonstrating the most consistent performance (Table 2). Slopes did not differ from unity only for comparisons against the loin cut surface, which was measured at the same anatomical site as the ultrasound scans; this was the case in three of the four B-mode combinations (Python-vertical, $\beta = 1.058$, $p = 0.083$; Python-parallel, $\beta = 0.983$, $p = 0.588$; ImageJ-vertical, $\beta = 0.989$, $p = 0.753$), whereas the ImageJ-parallel combination deviated slightly from unity ($\beta = 0.919$, $p = 0.014$). For all other references, slopes were significantly below unity ($p < 0.001$), indicating that B-mode values were not equivalent to those references.

In regressions based on hot carcass measurements, the vertical analysis yielded regression coefficients (β) ranging from 0.572 to 0.611, with R^2 values of 0.971–0.973 and RMSE values of 0.266–0.274. In contrast, the parallel analysis showed regression coefficients (β) of 0.558–0.598, R^2 values of 0.947–0.958, and higher RMSE values of 0.332–0.348, indicating greater prediction error than the vertical analysis. A similar pattern was observed under cold carcass

conditions. Notably, the combination of Python-based vertical analysis and cold carcass measurements showed the smallest prediction error ($\beta = 0.767$, $R^2 = 0.984$, $\text{RMSE} = 0.207$), although the slope differed significantly from unity ($p < 0.001$), whereas the parallel analysis exhibited relatively higher RMSE values (0.293–0.316).

Pearson correlation coefficients were also higher for the vertical analysis, indicating that B-mode vertical scans more accurately reflect carcass measurements. Consistent results were observed when compared with VCS2000 and loin cross-section measurements, where the vertical analysis showed lower RMSE than the parallel analysis. Specifically, the Python-based vertical analysis yielded $\beta = 0.821$, $R^2 = 0.972$, and $\text{RMSE} = 0.277$ when compared with VCS2000 measurements. When compared with loin cross-section measurements, it showed $\beta = 1.058$, $R^2 = 0.973$, and $\text{RMSE} = 0.274$, and the slope did not differ significantly from unity ($p = 0.083$), indicating equivalence with the reference measured at the same anatomical site. In contrast, the parallel analysis generally exhibited larger prediction errors, with the ImageJ-based parallel analysis showing the largest errors.

Evaluation of Mean Bias in B-mode Measurements

Paired *t*-tests were performed to assess whether B-mode ultrasound measurements systematically over- or underestimated the reference values (Table 3). The mean differences presented in Table 3 were calculated as the difference between the reference values and B-mode measurements (i.e., $\text{Reference} - \text{B-mode}$). Therefore, positive values indicate that B-mode measurements underestimated the reference values, whereas negative values indicate overestimation by B-mode.

When compared with direct measurements from the loin cross-section, the mean difference for the Python-based vertical analysis was -0.118cm , which was significantly different from zero ($p = 0.017$), indicating a tendency for overestimation relative to the reference. In contrast,

the mean difference for the parallel analysis (-0.037cm) was not statistically significant ($p = 0.451$). For the ImageJ-based analysis, neither the vertical (-0.025cm , $p = 0.642$) nor the parallel (0.062cm , $p = 0.240$) measurements showed significant differences compared with the loin cross-section reference.

In comparison with the field-based automated reference (VCS2000), both Python- and ImageJ-based analyses exhibited consistent and significant mean biases in both vertical and parallel orientations. The mean differences for the Python-based vertical and parallel analyses were 0.312cm and 0.314cm , respectively (both $p < 0.001$). Similarly, significant differences ranging from 0.411 to 0.415cm were observed in the ImageJ-based analyses (all $p < 0.001$), indicating that B-mode measurements were generally lower than those obtained from VCS2000.

More pronounced biases were observed when compared with carcass measurements. Under hot carcass conditions, the mean differences for B-mode measurements were 0.978 – 0.997cm for the Python-based analysis and 1.077 – 1.098cm for the ImageJ-based analysis, all of which were statistically significant ($p < 0.001$). Similarly, under cold carcass conditions, the mean differences ranged from 0.465 to 0.548cm for Python and from 0.561 to 0.648cm for ImageJ, with all differences being statistically significant ($p < 0.001$). These results indicate a systematic underestimation of carcass measurements by B-mode ultrasound.

Finally, comparisons between scanning orientations showed no significant differences, with mean differences of 0.054cm ($p = 0.063$) for the Python-based analysis and -0.002cm ($p = 0.963$) for the ImageJ-based analysis. In summary, only minor mean differences were observed for certain combinations when compared with the loin cross-section reference, whereas significant mean biases were consistently detected across all B-mode combinations when compared with VCS2000 and other carcass-based references. These findings suggest that calibration of B-mode measurements is necessary for direct comparison and integration with carcass-based reference data.

Relationship between A-mode and B-mode Measurements

Simple linear regression analyses were performed to evaluate the feasibility of converting A-mode measurements to the B-mode scale using the mean A-mode values as the independent variable and B-mode measurements as the dependent variable (Fig. 3). Across all four combinations (Python/ImageJ $\times$ vertical/parallel), strong linear relationships were observed between A-mode and B-mode measurements. The conversion coefficients were estimated with narrow 95% confidence intervals (1.205–1.330, 1.112–1.255, 1.116–1.251, and 1.036–1.179; Table 4), indicating that A-mode values can be converted to the B-mode scale using a single coefficient.

Among the models, the Python-based vertical analysis showed the lowest prediction error (RMSE = 0.233), while the ImageJ-based vertical analysis performed comparably (RMSE = 0.253). In contrast, models based on parallel scans exhibited higher prediction errors (RMSE $\approx$ 0.36) compared with the vertical models. Overall, the estimated regression coefficients ranged from 1.107 to 1.268, indicating that B-mode measurements tended to yield higher values than A-mode measurements for the same input (Table 4).

Predictive Performance of A-mode for Carcass Measurements

Field applicability was evaluated by assessing whether A-mode measurements alone could directly predict carcass reference values. Simple linear regression analyses were performed using A-mode values as the independent variable and each reference value (loin cross-section measurement, VCS2000, cold carcass, and hot carcass) as the dependent variable (Table 5). Regression slopes ranged from 1.199 to 1.879 and were significantly greater than unity in all cases ($p < 0.001$), indicating systematic underestimation by A-mode that increased with the physiological distance between the *in vivo* measurement and the reference.

In particular, the smallest prediction errors were observed for VCS2000 (RMSE=0.341) and the loin cross-section measurement (RMSE=0.301). Prediction errors tended to increase under carcass conditions, with RMSE values of 0.461 for cold carcass and 0.600 for hot carcass measurements. Overall, these results indicate that raw A-mode values cannot be treated as equivalent to carcass reference values, but that they can be converted to the reference scale by applying the single calibration coefficient estimated for each reference (Table 5).

Statistical Comparison between A-mode Measurements and Carcass Reference Values

Paired *t*-tests were performed to evaluate whether A-mode measurements provide values comparable to carcass reference values for both raw A-mode measurements and regression-based calibrated (predicted) values (Table 8). Prior to calibration, significant mean differences were observed for all reference values (loin cross-section, VCS2000, cold carcass, and hot carcass; all $p < 0.001$). The mean differences presented here were calculated as the difference between the reference values and A-mode measurements (i.e., Reference – A-mode); positive values therefore indicate underestimation by A-mode. Specifically, the mean differences were 0.276cm for the loin cross-section, 0.599cm for VCS2000, 0.901cm for the cold carcass, and 1.245cm for the hot carcass, indicating that A-mode measurements generally underestimated the reference values.

In contrast, the differences between the predicted values and the reference values were no longer statistically significant across all references after applying regression-based calibration (loin cross-section: $p = 0.528$; VCS2000: $p = 0.298$; cold carcass: $p = 0.201$; hot carcass: $p = 0.142$). The mean differences were substantially reduced to 0.026cm, 0.043cm, 0.072cm, and 0.106cm for the loin cross-section, VCS2000, cold carcass, and hot carcass, respectively. These results indicate that although raw A-mode measurements significantly differed from the

reference values, the application of regression-based calibration eliminated these differences, rendering A-mode data compatible with carcass-based reference scales.

The robustness of these calibration models was further examined by leave-one-out cross-validation (Table 8). Cross-validated prediction errors exceeded the corresponding fitted errors by only 0.64–1.08%, indicating that the single-coefficient models were not overfitted to the present dataset. To verify equivalence without circularity, the slope test was repeated after re-estimating each calibration coefficient with the corresponding observation excluded; slopes ranged from 0.997 to 0.999 and all 95% confidence intervals included unity ($p = 0.924$ – 0.961). Lin's concordance correlation coefficient increased after calibration for all references, with the largest relative increase observed for hot carcass (0.092 to 0.483), and the bias correction factor approached unity in all comparisons (0.951–0.997). White's specification test provided no evidence of heteroscedasticity ($p = 0.369$ – 0.860), supporting the adequacy of the linear specification across the observed range. Bland–Altman analysis showed that mean bias was reduced by 90.6–92.8% after calibration, although the limits of agreement remained relatively wide (1.17–2.31cm; Fig. 4).

Compatibility between A-mode and B-mode Measurements

Following the decision framework illustrated in Figure 1, this study evaluated whether A-mode and B-mode ultrasound systems are mutually compatible under a unified carcass-based reference scale. To this end, three sequential steps were assessed: (1) the degree to which B-mode measurements reflect carcass reference values (B→Reference), (2) the feasibility of converting A-mode measurements to the B-mode scale (A→B), and (3) the direct predictive performance of A-mode for carcass reference values (A→Reference). At each step, not only linear relationships but also the statistical significance of mean bias was evaluated using paired t -tests.

The results showed that B-mode (Python-based vertical analysis) values deviated systematically from the identity line with carcass reference values. In addition, significant mean differences were observed between raw B-mode measurements and both VCS2000 and carcass measurements ($p < 0.001$), indicating the presence of systematic bias that prevents direct integration without calibration. The conversion of A-mode measurements to the B-mode scale was described by a stable single coefficient ($\beta = 1.107\text{--}1.268$) with low prediction error across all four analytical combinations, confirming that A-mode data can be reliably transformed into the more precise B-mode scale. Furthermore, evaluation of direct prediction models from A-mode to carcass reference values revealed that raw A-mode data showed significant differences from all reference values ($p < 0.001$), whereas these differences became non-significant after applying regression-based calibration ($p = 0.142\text{--}0.528$).

This indicates that the calibration models effectively removed systematic bias associated with device-specific characteristics. Overall, although A-mode and B-mode measurements are not directly compatible in their raw forms due to differences in measurement principles, the application of the conversion and calibration models developed in this study enables their integration without significant error on a carcass-based reference scale. Therefore, the relationship between the two systems can be classified as “compatible after calibration”, indicating conditional compatibility contingent upon appropriate standardization (Table 6).

Accuracy Evaluation and Optimization of the B-mode Ultrasound Measurement Protocol

This study evaluated the effects of probe scanning orientation and image analysis tools on the accuracy of predicting carcass reference values using B-mode ultrasound, with the aim of establishing an optimal measurement protocol (Table 7). Analysis based on the Python algorithm showed that scanning perpendicular to the spine consistently yielded lower

prediction error than parallel scanning. In regressions against loin cut surface measurements, vertical scanning gave the lowest RMSE (0.274) and mean absolute error (0.209cm), whereas parallel scanning showed RMSE 0.356 and MAE 0.264cm, corresponding to approximately 1.3-fold greater prediction error. The corresponding R^2 values (0.973 vs. 0.951) followed the same ordering but are reported as descriptive summaries only.

Notably, differences between the two scanning orientations were more pronounced in terms of error stability. Analysis of absolute errors revealed that the maximum error for vertical scanning was controlled at approximately 0.83cm, whereas parallel scanning reached up to 1.18cm. This discrepancy is likely attributable to increased variability in parallel scans, where alignment with muscle fiber direction may introduce noise from intramuscular fat or irregular fascia structures.

A similar pattern was observed when using the general-purpose image analysis tool ImageJ. Under ImageJ analysis, the mean absolute error (MAE) for vertical scanning was 0.215cm, which was lower than that of parallel scanning (0.276cm), and the maximum error was also substantially lower for vertical scanning (0.96cm) compared with parallel scanning (1.27cm).

Although the overall error levels of ImageJ (manual measurement) were slightly higher than those of the Python-based automated analysis, the trend that “vertical scanning results in lower error and reduced risk of outliers compared with parallel scanning” was consistently observed regardless of the analysis tool. Paired t -tests indicated no statistically significant differences in mean error between scanning orientations ($p > 0.05$). However, vertical scanning was confirmed to be a more stable protocol due to its lower maximum error from a risk management perspective.

Discussion

Systematic Differences between *In Vivo* Ultrasound and Carcass Measurements and Their Physical Mechanisms

It should be emphasized that the analytical question addressed in this study was whether two measurement systems yield equivalent values, rather than how strongly one variable explains the other. Under this framing, the relevant criterion is whether the regression slope equals unity, and the coefficient of determination is not an appropriate measure of measurement agreement.

The results of this study showed that *in vivo* B-mode ultrasound measurements were generally comparable to direct measurements obtained from the loin cross-section at the same anatomical location, with three of the four analytical combinations showing no significant mean difference in paired *t*-tests ($p > 0.05$; Table 3). The Python-based vertical analysis showed a small but significant overestimation of 0.118cm ($p = 0.017$), which nonetheless remained within the resolution limits of the measurement systems. However, despite being derived from the same individuals, backfat thickness measured under hot and cold carcass conditions showed significant increases compared with *in vivo* measurements, with maximum differences of 1.09cm and 0.65cm, respectively. These discrepancies are interpreted not as simple measurement errors but as systematic bias arising from changes in tissue properties and physical deformation during the harvest process. First, the effect of carcass suspension should be considered, as it induces longitudinal stretching of the body axis and gravitational redistribution of adipose tissue, leading to localized thickening due to tissue redistribution and structural deformation. In addition, post-mortem loss of muscle tone may also contribute to these discrepancies. Following harvest, relaxation of muscle tissues may reduce structural support surrounding the backfat layer, allowing slight deformation or redistribution of adjacent tissues, which could partially influence the measured backfat thickness.

According to Sather et al. [16], the body axis is horizontally aligned under live conditions, whereas the orientation of gravitational force changes when carcasses are suspended by the

hind legs after harvest. Under these conditions, vertical loading leads to compression of the backfat layer along the longitudinal axis, resulting in an apparent increase in thickness in the transverse plane, a phenomenon commonly referred to as “bunching.” The increase in carcass measurements observed in this study is consistent with this compression effect and agrees with previous reports indicating that carcass measurements tend to be greater than *in vivo* measurements [17, 18]. Notably, the magnitude of the discrepancy between B-mode (*in vivo*) and carcass measurements in this study (approximately $\geq 1\text{cm}$) was greater than that reported in earlier studies (approximately 0.1–0.4cm). This difference is likely attributable to the increased market weight of modern finishing pigs. In earlier studies, the harvest weights were approximately 80kg [18], 90–92kg [16], and 102.6kg [17]. In contrast, the animals used in the present study reflected current industry trends toward heavier weights, with a mean live weight of $120.6 \pm 4.723\text{kg}$ (range: 113–132kg).

Second, temperature-dependent changes in the physical properties of adipose tissue must be considered. Porcine adipose tissue, composed of a mixture of fatty acids, exhibits complex melting behavior within the temperature range of approximately 25–50°C [3]. Backfat exists in a partially melted or semi-solid state with a loosened crystalline structure at body temperature (approximately 39°C) and under hot carcass conditions immediately after harvest, resulting in high fluidity. Consequently, it is more susceptible to deformation under gravitational loading during suspension. In contrast, the reduction in discrepancy after 24h of chilling (cold carcass) can be attributed to fat crystallization and an increase in solid fat content, which enhance tissue stiffness and thereby mitigate compression and redistribution effects.

The additional agreement analyses conducted in this study also clarify the nature of the compatibility that the calibration models provide. Calibration removes systematic bias but does not reduce random error; when the scaling coefficient exceeds unity, random error is rescaled proportionally together with the systematic component. Consequently, the limits of agreement

widened after calibration (for example, from 1.75 to 2.31cm for hot carcass), even though mean bias was reduced by more than 90%. This pattern is also reflected in the decomposition of Lin's concordance correlation coefficient: the bias correction factor approached unity in all comparisons (0.951–0.997), indicating that the residual disagreement after calibration is attributable to precision rather than to systematic offset. The compatibility demonstrated here should therefore be understood as agreement at the population level rather than individual-level interchangeability. Calibrated A-mode values are suitable for integrating phenotypic records collected with different devices in genetic evaluation, where estimates are derived from many animals, but are not intended to substitute for carcass grading of individual animals.

Limitations of the Study

Although this study provides meaningful insights by elucidating the compatibility between A-mode and B-mode ultrasound systems and proposing calibration models that account for underlying physical mechanisms, several limitations should be considered when interpreting the results and applying them in field conditions.

There are limitations related to sample size and breed. This study was conducted on 69 Yorkshire pigs. Although the calibration coefficients were estimated with narrow confidence intervals and showed stable performance under cross-validation, the applicability of the derived calibration equations to other breeds with different genetic backgrounds, such as Duroc or Landrace, or to differences in sex (female, male, or castrated), may be limited. In addition, commercial crossbred populations and differences in production environments, including feeding systems and management conditions, may further influence the applicability of the proposed calibration equations. Therefore, future studies should validate the generalizability of the proposed models using larger populations and a wider range of breeds.

Operator variability and potential nonlinearity should be considered. In the present study, all ultrasound measurements and image analyses were intentionally performed by a single trained operator to minimize operator-induced variability and isolate device-related differences during inter-system compatibility assessment. Although this design improved internal consistency for device comparison, future studies involving multiple operators under practical field conditions are warranted to further validate the robustness and field applicability of the proposed calibration models. In field conditions, measurements are typically performed by multiple operators rather than a single trained individual, and differences in operator skill may result in variation in probe pressure, leading to physical deformation (compression) of the fat layer and additional measurement error [9, 16, 19, 20].

Moreover, operator-dependent behavioral bias may also contribute to systematic underestimation in the case of A-mode measurements. When relatively large values are obtained, operators may question their validity and repeat measurements multiple times, often recording lower values once they are observed. This tendency may lead to preferential selection of smaller readings, thereby introducing a downward bias in A-mode measurements [18, 20].

Because vertical scanning was introduced during the study period, vertical and parallel images were available for 35 and 59 animals, respectively ($n = 31$ and 54 after pairwise deletion for the comparisons in Table 7), and the comparison between scanning orientations in Table 7 was therefore based on partially different subsets of animals. This comparison should be interpreted with caution, as differences in subset composition may contribute to the observed differences in error metrics.

The nature of the agreement achieved by calibration should also be considered. Although calibration effectively removed systematic bias, the 95% limits of agreement after calibration ranged from 1.17 to 2.31cm, corresponding to 70–91% of the mean backfat thickness of the respective reference measurements. The calibrated values are therefore appropriate for

adjusting systematic differences between devices at the population level, such as when integrating phenotypic records for genetic evaluation, but are not intended to replace carcass grading of individual animals. In addition, the calibration equations should be applied within the observed range of A-mode measurements (0.70–2.57cm); a small number of animals at the upper end of this range showed deviations exceeding 1cm after calibration, and extrapolation beyond the observed range is not recommended.

Furthermore, the predictive performance of linear models may be limited at extreme levels of backfat thickness, where nonlinear patterns and increased measurement variability are more likely to occur. Therefore, further studies are warranted to validate the robustness of the proposed models under practical conditions.

Differences in the anatomical definition of the measurements should also be noted. A-mode and B-mode values represent a single site between the 10th and 11th thoracic vertebrae, whereas hot and cold carcass values represent the mean of three sites (shoulder, 10th–11th thoracic vertebrae, and hip region). Backfat thickness varies considerably among anatomical sites, and this definitional difference is expected to contribute to the discrepancies observed between *in vivo* and carcass measurements, in addition to the physical mechanisms discussed above. Consistent with this interpretation, slopes that did not differ from unity were observed only for comparisons against the loin cut surface, which was measured at the same anatomical site as the ultrasound scans (three of the four B-mode combinations), and the bias correction factor decreased progressively from the loin cut surface (0.997) to the hot carcass (0.951).

The proposed calibration framework may provide a practical basis for integrating heterogeneous backfat thickness records obtained from different ultrasound systems in national breeding programs and commercial farm settings. Such standardization could improve phenotypic consistency across measurement platforms and enhance the reliability of genetic evaluation and selection decisions. However, large-scale validation across multiple breeds,

crossbred populations, and production environments is required before routine field implementation.

References

1. Boland M, Foster K, Schinckel A, Wagner J, Chen W, Berg E, Forrest J. Alternative pork carcass evaluation techniques: I. Differences in predictions of value. *J Anim Sci.* 1995;73(3):637-644. doi: 10.2527/1995.733637x
2. Suzuki K, Irie M, Kadowaki H, Shibata T, Kumagai M, Nishida A. Genetic parameter estimates of meat quality traits in Duroc pigs selected for average daily gain, longissimus muscle area, backfat thickness, and intramuscular fat content. *J Anim Sci.* 2005;83(9):2058-2065. doi: 10.2527/2005.8392058x
3. Wood JD, Enser M, Fisher AV, Nute GR, Sheard PR, Richardson RI, Hughes SI, Whittington FM. Fat deposition, fatty acid composition and meat quality: A review. *Meat Sci.* 2008;78(4):343-358. doi: 10.1016/j.meatsci.2007.07.019
4. Cross HR, Belk KE. Objective measurements of carcass and meat quality. *Meat Sci.* 1994;36(1-2):191-202. doi: 10.1016/0309-1740(94)90041-8
5. Choi JG, Cho CI, Choi IS, Lee SS, Choi TJ, Cho KH, Park BH, Choy YH. Genetic parameter estimation in seedstock swine population for growth performances. *Asian-Australas J Anim Sci.* 2013;26(4):470-475. doi: 10.5713/ajas.2012.12454
6. Herring WO, Miller DC, Bertrand JK, Benyshek LL. Evaluation of machine, technician, and interpreter effects on ultrasonic measures of backfat and longissimus muscle area in beef cattle. *J Anim Sci.* 1994;72(9):2216-2226. doi: 10.2527/1994.7292216x
7. Fortin A, Tong AKW, Robertson WM. Evaluation of three ultrasound instruments, CVT-2, UltraFom 300 and AutoFom for predicting salable meat yield and weight of lean in the primals of pork carcasses. *Meat Sci.* 2004;68(4):537-549. doi: 10.1016/j.meatsci.2004.05.006
8. Wagner DR, Thompson BJ, Anderson DA, Schwartz S. A-mode and B-mode ultrasound measurement of fat thickness: A cadaver validation study. *Eur J Clin Nutr.* 2019;73:518-523. <https://doi.org/10.1038/s41430-018-0085-2>
9. Moeller SJ, Christian LL. Evaluation of the accuracy of real-time ultrasonic measurements of backfat and loin muscle area in swine using multiple statistical analysis procedures. *J Anim Sci.* 1998;76(10):2503-2514. doi: 10.2527/1998.76102503x
10. Lucas D, Brun A, Gispert M, Carabús A, Soler J, Tibau J, Font-i-Furnols M. Relationship between pig carcass characteristics measured in live pigs or carcasses with Piglog, Fat-o-Meat'er and computed tomography. *Livest Sci.* 2017;197:88-95. doi:10.1016/j.livsci.2017.01.010

- 579 11. Newcom DW, Baas TJ, Lampe JF. Prediction of intramuscular fat percentage in live
580 swine using real-time ultrasound. *J Anim Sci.* 2002;80(12):3046-3052. doi:
581 10.2527/2002.80123046x
- 582 12. Scholz AM, Bünger L, Kongsro J, Baulain U, Mitchell AD. Non-invasive methods for
583 the determination of body and carcass composition in livestock: dual-energy X-ray
584 absorptiometry, computed tomography, magnetic resonance imaging and ultrasound:
585 invited review. *Animal.* 2015;9:1250-1264. doi:10.1017/S1751731115000336
- 586 13. Houghton PL, Turlington LM. Application of ultrasound for feeding and finishing
587 animals: a review. *J Anim Sci.* 1992;70(3):930-941. doi: 10.2527/1992.703930x
- 588 14. Cisneros F, Ellis M, Miller KD, Novakofski J, Wilson ER, McKeith FK. Comparison of
589 transverse and longitudinal real-time ultrasound scans for prediction of lean cut yields
590 and fat-free lean content in live pigs. *J Anim Sci.* 1996;74(11):2566-2576. doi:
591 10.2527/1996.74112566x
- 592 15. Kim JW, Kang MH, Choi SW, Dang CG, Lee SM, Park B, Yang HS, Oh SH. PSVII-18:
593 Standardization method for ultrasound device measurements of pig loin area:
594 Compatibility study between A-mode and B-mode devices. *J Anim Sci.* 2025;103
595 (Supplement_3):663-664. doi: 10.1093/jas/skaf300.749
- 596 16. Sather AP, Fredeen HT, Martin AH. Live animal evaluation of two ultrasonic probes as
597 estimators of subcutaneous backfat and carcass composition in pigs. *Can J Anim Sci.*
598 1982;62:943-949. doi: 10.4141/cjas82-114
- 599 17. Turlington LM. Live animal evaluation of carcass traits for swine and sheep using real-
600 time ultrasound [Master's thesis]. Manhattan (KS): Kansas State University; 1989
- 601 18. McPhee CP, Daniels LJ. Effects of genotype, diet and sex on backfat depth in pigs
602 measured physically at different carcass sites and ultrasonically at different liveweights.
603 *Aust J Exp Agric.* 1991;31:761-764. doi: 10.1071/EA9910761
- 604 19. Magowan E, McCann MEE. A comparison of pig backfat measurements using ultrasonic
605 and optical instruments. *Livest Sci.* 2006;103:116-123. doi:10.1016/j.livsci.2006.02.002
- 606 20. Peppmeier ZC, Howard JT, Knauer MT, Leonard SM. Estimating backfat depth, loin
607 depth, and intramuscular fat percentage from ultrasound images in swine. *Animal.*
608 2023;17:100969. doi: 10.1016/j.animal.2023.100969

609 Table 1. Descriptive statistics of ultrasound and carcass measurements for Yorkshire pigs

Category	Variables	n	Mean±SD	Min	Max
General Traits	Carcass Weight(kg)	69	90.02±4.047	81.6	98.8
	Fresh Weight(kg)	69	120.6±4.723	113	132
Ultrasound (Live)	A-mode(cm)	69	1.297±0.349	0.7	2.567
	B-mode Parallel Python(cm)	59	1.576±0.282	1.091	2.327
	B-mode Vertical Python(cm)	35	1.611±0.291	1.039	2.331
	B-mode Parallel ImageJ(cm)	59	1.475±0.289	0.961	2.216
	B-mode Vertical ImageJ(cm)	35	1.512±0.255	1.035	2.096
	Hot Carcass(cm)	69	2.542±0.505	1.5	3.8
Carcass (Post-mortem)	Cold Carcass(cm)	66	2.191±0.450	1.3	3.6
	VCS2000(cm)	69	1.896±0.453	0.9	3.3
	Loin Cut Surface(cm)	54	1.531±0.391	0.877	2.663

610

611

Table 2. Summary of regression and correlation statistics for B-mode ultrasound measurements regressed on reference backfat thickness measurements

B-mode Method	Reference	n	Slope (β)	p ($\beta=1$)	R^2	RMSE	Correlation (r)
Python (Vertical)	Hot Carcass	35	0.611	<0.001	0.973	0.274	0.613
	Cold Carcass	32	0.767	<0.001	0.984	0.207	0.771
	VCS2000	35	0.821	<0.001	0.972	0.277	0.63
	Loin Cut Surface	31	1.058	0.083	0.973	0.274	0.645
Python (Parallel)	Hot Carcass	59	0.598	<0.001	0.958	0.332	0.34
	Cold Carcass	56	0.727	<0.001	0.967	0.293	0.482
	VCS2000	59	0.806	<0.001	0.953	0.348	0.422
	Loin Cut Surface	54	0.983	0.588	0.951	0.356	0.482
ImageJ (Vertical)	Hot Carcass	35	0.572	<0.001	0.971	0.266	0.562
	Cold Carcass	32	0.715	<0.001	0.973	0.255	0.575
	VCS2000	35	0.766	<0.001	0.964	0.293	0.499
	Loin Cut Surface	31	0.989	0.753	0.962	0.303	0.468
ImageJ (Parallel)	Hot Carcass	59	0.558	<0.001	0.947	0.348	0.241
	Cold Carcass	56	0.679	<0.001	0.956	0.316	0.372
	VCS2000	59	0.753	<0.001	0.945	0.356	0.351
	Loin Cut Surface	54	0.919	0.014	0.941	0.366	0.395

R^2 was obtained from models fitted through the origin and is therefore computed relative to the uncorrected total sum of squares; it is not directly comparable to r^2 and does not indicate measurement agreement. p ($\beta=1$) refers to the test of the null hypothesis that the slope equals unity, corresponding to the identity line ($y = x$).

619 Table 3. Mean differences (cm) and paired *t*-test *p*-values (in parentheses) for assessing systematic
620 bias of B-mode backfat thickness measurements

Reference	Vertical		Parallel	
	Python	ImageJ	Python	ImageJ
Loin Cut Surface	-0.118(0.017)	-0.025(0.642)	-0.037(0.451)	0.062(0.24)
VCS2000	0.312(<0.001)	0.411(<0.001)	0.314(<0.001)	0.415(<0.001)
Hot Carcass	0.978(<0.001)	1.077(<0.001)	0.997(<0.001)	1.098(<0.001)
Cold Carcass	0.465(<0.001)	0.561(<0.001)	0.548(<0.001)	0.648(<0.001)

621 Sample sizes differ among combinations owing to pairwise complete cases (n = 31–59).

622

623 Table 4. Conversion equations from A-mode to B-mode backfat thickness

B-mode	n	Slope (β)	SE	95%CL	p ($\beta=1$)	R^2	RMSE
Python (Vertical)	35	1.268	0.031	1.205–1.33	<0.001	0.980	0.233
Python (Parallel)	59	1.183	0.036	1.112–1.255	<0.001	0.95	0.361
ImageJ (Vertical)	35	1.183	0.033	1.116–1.251	<0.001	0.974	0.253
ImageJ (Parallel)	59	1.107	0.036	1.036–1.179	0.004	0.943	0.362

624 R^2 was obtained from models fitted through the origin and is therefore computed relative to the
625 uncorrected total sum of squares; it is not directly comparable to r^2 and does not indicate
626 measurement agreement. p ($\beta=1$) refers to the test of the null hypothesis that the slope equals unity,
627 corresponding to the identity line ($y = x$).

628

629 Table 5. Direct prediction models from A-mode to carcass/reference backfat thickness

Reference (Y)	n	β	$p (\beta=1)$	R^2	RMSE	Correlation (r)
Loin Cut Surface	54	1.199	<0.001	0.964	0.301	0.697
VCS2000	69	1.429	<0.001	0.97	0.341	0.751
Cold carcass	66	1.643	<0.001	0.958	0.461	0.619
Hot carcass	69	1.879	<0.001	0.947	0.6	0.508

630 R^2 was obtained from models fitted through the origin and is therefore computed relative to the
631 uncorrected total sum of squares; it is not directly comparable to r^2 and does not indicate
632 measurement agreement. $p (\beta=1)$ refers to the test of the null hypothesis that the slope equals unity,
633 corresponding to the identity line ($y = x$).

634

635

636 Table 6. Final evaluation of compatibility between A-mode and B-mode ultrasound measurements
637 based on regression performance and systematic bias assessment

Evaluation Step	Relationship	Agreement Criterion	Paired t -test	Verdict
1. B-mode Accuracy	B-mode vs Reference ¹	Slope deviates from unity except for the loin cut surface ³	$p < 0.001$ (Bias Exists) ³	Requires Calibration
2. Inter-device Conversion	A-mode $\rightarrow$ B-mode	Stable single-coefficient conversion ($\beta = 1.107$ – 1.268)	-	Convertible(Stable Relationship)
3. Predictive Validity	A-mode vs Reference ¹	Slope deviates from unity ($\beta = 1.199$ – 1.879 , $p < 0.001$)	$p < 0.001$ (Bias Exists)	Incompatible(Raw Data)
	A-mode ² vs Reference	Slope does not differ from unity under cross-validation ($p = 0.924$ – 0.961)	$p > 0.05$ (Bias Removed)	Compatible(After Calibration)

¹Reference includes Loin Cut Surface, VCS2000, Cold Carcass, and Hot Carcass.

²Calibrated using the simple linear regression models derived in this study.

³For the slope test, p ($\beta=1$) was < 0.001 for VCS2000, cold carcass, and hot carcass in all four B-mode combinations, whereas slopes did not differ from unity for the loin cut surface in three of the four combinations (Python-vertical, Python-parallel, and ImageJ-vertical; $p = 0.083$ – 0.753). For the paired t -test, mean bias was significant for VCS2000, cold carcass, and hot carcass, and non-significant for the loin cut surface except in the Python-vertical combination ($p = 0.017$).

638

Table 7. Comparison of predictive performance and error metrics between vertical and parallel scanning protocols in B-mode ultrasound

Scanning Protocol	Software	RMSE	Mean Absolute Error (cm)	Max Error (cm)	R^2
Vertical	Python	0.274	0.209	0.828	0.973
	ImageJ	0.303	0.215	0.955	0.962
Parallel	Python	0.356	0.264	1.176	0.951
	ImageJ	0.366	0.276	1.269	0.941

Values were obtained from regressions of B-mode measurements on loin cut surface measurements ($n = 31$ for vertical and 54 for parallel). R^2 was obtained from models fitted through the origin and does not indicate measurement agreement.

645 Table 8. Cross-validated performance and agreement of the A-mode calibration models

Item	Loin Cut Surface	VCS2000	Cold carcass	Hot carcass
n	54	69	66	69
Mean difference before calibration (cm)	0.276	0.599	0.901	1.245
Paired <i>t</i> -test <i>p</i> , before calibration	<0.001	<0.001	<0.001	<0.001
Mean difference after calibration (cm)	0.026	0.043	0.072	0.106
Paired <i>t</i> -test <i>p</i> , after calibration	0.528	0.298	0.201	0.142
RMSE (LOOCV, cm)	0.303	0.343	0.466	0.607
Δ over fitted RMSE (%)	0.91	0.64	0.98	1.08
Slope (LOOCV)	0.999	0.999	0.998	0.997
95% CI	0.945–1.052	0.956–1.042	0.945–1.050	0.939–1.055
<i>p</i> ($\beta=1$)	0.961	0.956	0.934	0.924
CCC before calibration	0.521	0.344	0.169	0.092
CCC after calibration	0.695	0.745	0.600	0.483
C _b after calibration	0.997	0.991	0.970	0.951

646 Mean differences were calculated as reference minus A-mode; positive values indicate underestimation by
647 A-mode; in Fig. 4 the same differences are plotted in the opposite direction (A-mode minus reference).
648 RMSE (LOOCV) is the leave-one-out prediction error; Δ is its relative increase over the fitted RMSE

649 reported in Table 5. Slope (LOOCV) was obtained by applying to each observation a calibration coefficient
650 re-estimated with that observation excluded; p ($\beta=1$) refers to the test of the null hypothesis that the slope
651 equals unity, corresponding to the identity line ($y = x$). Using the full-data calibration coefficients,
652 corresponding slopes were 0.9998–1.0003 ($p = 0.990$ – 0.999). CCC, Lin's concordance correlation
653 coefficient; C_b , bias correction factor. White's specification test provided no evidence of heteroscedasticity
654 in any comparison ($p = 0.369$ – 0.860).

Figure 1. Flowchart of the decision-making process for evaluating compatibility between A-mode and B-mode ultrasound devices

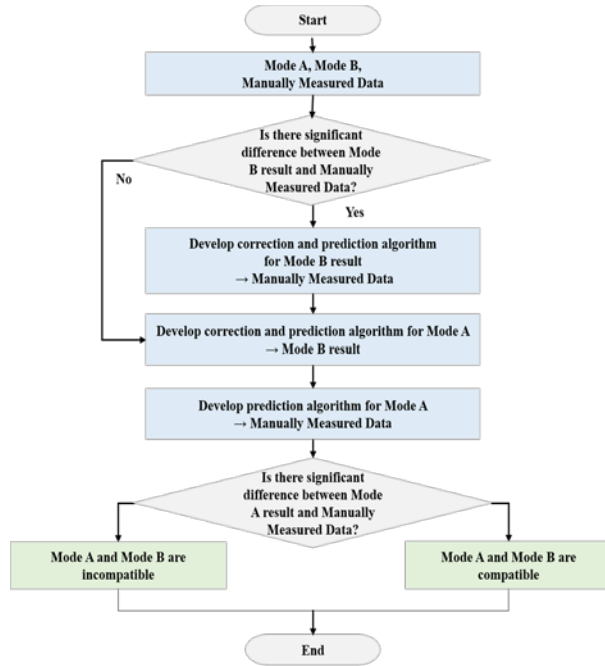
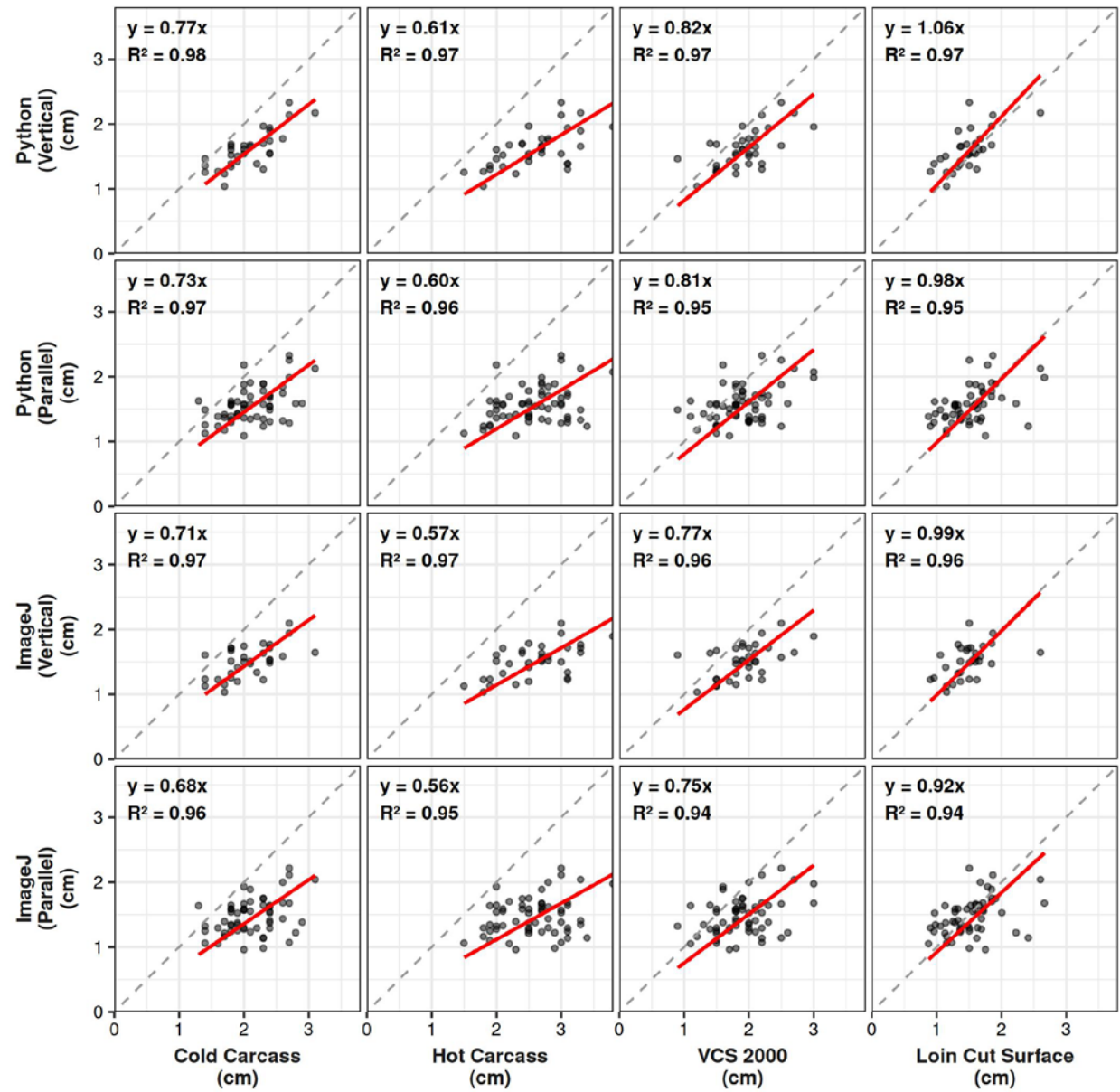
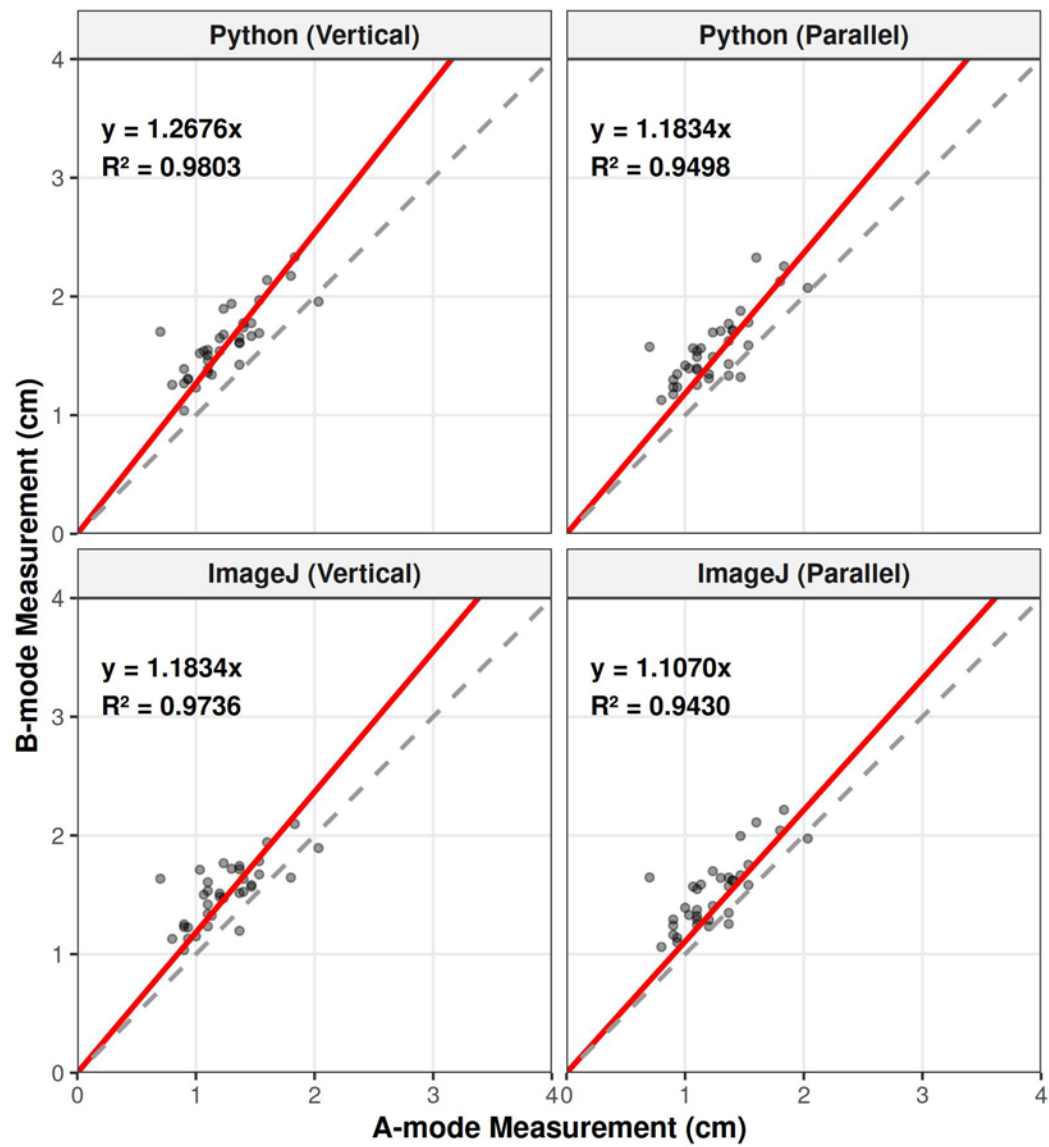


Figure 2. Faceted scatter plots showing relationships between B-mode-derived backfat thickness and reference measurements (Cold/Hot carcass, VCS2000, and Loin Cut Surface), with fitted regression lines and identity line ($y=x$)



665 Figure 3. Relationships between A-mode and B-mode backfat thickness measurements and A-to-
666 B conversion models



667

668

669

Figure 4. Bland–Altman plots comparing A-mode measurements with reference backfat thickness measurements before (left) and after (right) calibration. Solid red lines indicate mean bias, dotted red lines indicate 95% limits of agreement, and dashed grey lines indicate zero difference. Differences are plotted as A-mode minus reference, so that negative values indicate underestimation by A-mode; the corresponding values in Table 8 are presented as reference minus A-mode.

