AI-Driven Prediction of Cattle Body Weight Using Morphometric Measurements

Bhavya Shah
Department of Computer Engineering and Technology
Dr. Vishwanath Karad MIT World Peace University
Pune, Maharastra, India
bhavyashah16112004bs@gmail.com

Lakshya Upadhyaya
Department of Computer Engineering and Technology
Dr. Vishwanath Karad MIT World Peace University
Pune, Maharastra, India
lakshya.upadhyaya05@gmail.com

Samarth Patel
Department of Computer Engineering and Technology
Dr. Vishwanath Karad MIT World Peace University
Pune, Maharastra, India
samarthpatel23104@gmail.com

Gautam Sharma
Department of Computer Engineering and Technology
Dr. Vishwanath Karad MIT World Peace University
Pune, Maharastra, India
sharma.gautam0905@gmail.com

Shilpa Sonawani
Department of Computer Engineering and Technology
Dr. Vishwanath Karad MIT World Peace University
Pune, Maharastra, India
Shilpa.sonawani@mitwpu.edu.in

Abstract— **The adoption of Artificial Intelligence (AI) in managing livestock has transformed the process of animal evaluation through the introduction of non-invasive methods of data-driven and reliable evaluation. The present paper is an intelligent predictive model of body weight and associated biological characteristics of cattle through morphometric measurements. The suggested system is a combination of the Shaeffer's empirical formula and advanced machine learning (ML) methods to improve the interpretability and accuracy. The physical data is composed of the wither height, rump height, body length, chest depth, girth measurements, and bone diameters. The extensive preprocessing was done to manage the noise, equalize the distribution of features, and enhance consistency in the model.**

The feature engineering added variables that are domain specific such as Heart Girth Square and Height-Girth Index that express non-linear morphometric associations and volumetric development patterns. Comparative analysis, of both multiple regression and ensemble models, showed that the empirical equations together with the ML learning show better prediction performance. The article demonstrates the promise of AI-based morphometric analytics as a paradigm-shift in the ongoing agricultural system.

Keywords—Cattle analysis, machine learning, morphometric data, weight prediction, livestock management.

I. INTRODUCTION

The field of agricultural production and livestock care is no exception, as AI and ML are transforming a number of industries. Cattle rearing in the world economy is a crucial agricultural industry and it is a source of meat, milk, and other valuable products. Effective livestock management involves effective monitoring of the health and body condition of the animals and the productivity indices. One of these is the weight of the body, which is a very important parameter which directly determines the feeding strategies, breeding programs, and market assessment. Conventionally, weighing scales or visual estimation are some of the methods that are used to determine cattle weight; these two methods are usually laborious, animal stressful and highly vulnerable to human errors.

The latest developments in data analytics and machine learning have allowed automatic and data-driven solutions to address these shortcomings. ML models based on morphometric measurements can predict the body weight of cattle or classify cattle breeds with a high precision by using height, girth, body length and bone diameters. The methods are also able to minimize manual labor in addition to offering reliable, non-invasive, and affordable options to farmers and researchers.

This paper seeks to build a machine learning pipeline that will be able to predict the weight of cattle given physical measurements as one of the input features. The data is further divided into several morphometric parameters, such as wither height, rump height, body length, chest depth, abdominal girth and bone diameters. To enhance the quality of data and guarantee the reliability of the models, the preprocessing methods of normalization, feature scaling, and outlier removal were used. The implementation and comparison of different algorithms, such as linear regression, random forest, and ensemble learning models, were conducted based on such metrics as R2 score, mean square error (MSE), and mean absolute error (MAE).

The suggested system offers an effective and scalable system of estimating cattle weights in the actual world, especially in rural or resource-restricted settings. This study facilitates smarter livestock management and will increase the productivity of livestock by applying predictive analytics to livestock management, leading to the sustainable agriculture concept on a larger scale. Here, the related literature, methodology, experimental results, and main conclusions made in the study are presented in the following sections.

II. LITERATURE SURVEY

Proper body weight estimation of cattle improves herd feeding, health management and economic efficiency but standard weighing is cumbersome, expensive and may cause stress to animals. This is the reason, researchers, have resorted to morphometric measures alongside machine learning to provide low impact, fast predictions [1-2].

A. Linear and Regression-Based Models:

Linear Regression is usually employed as a benchmark. Pearson correlation coefficient of as low as 0.12-0.48 and high RMSE values (60-309.8 kg) in a 3D-image study demonstrate that simple linear, ridge and lasso models are the least effective models in predicting a variable (as shown by the low coefficients and high RMSE) particularly when using small sets of jersey (2). An experiment of dairy cows with chest, abdominal perimeter and rump width found single variable R2 of almost 0.78, with multiple-linear (CP + AP + RW) model had almost 0.91 R2 indicating usefulness of incorporation of key predictors [1]. Linear regression demonstrated the most overall effectiveness (MAE 17.8kg, R2 = 0.83) in one type of species weight analysis (livestock, varied species) that used chest circumference, maclock width and chest depth as part of the important features [3].

B. Tree-Based Ensemble Models:

In various studies, tree ensembles are always more successful than linear models. CatBoost, AdaBoost and Random Forest was identified as the 3D-image work that provided the best R2, lowest RMSE and least variance with CatBoost as the best among them [2]. According to the dairy-cattle paper, even though it is merely the promise of Randon Forest it suggests the ensemble methods as a potential next step [1]. Gradient Boosting (MAE 20.1kg, R2 0.79) Lagged behind linear regression, whereas Random Forest slightly outperformed the former (MAE 21.1kg, R2 0.79) in the livestock-weight paper Anjar Setiwan [3]. The study of pig weight indicates that a deep learning model as opposed to tree methods was an indicator of change towards more complex architecture.

C. Neural Networks and Sophisticated Models:

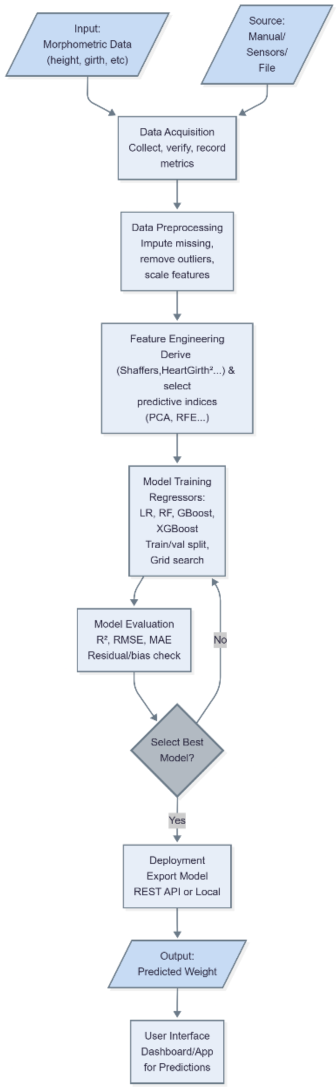
The best gains are offered by artificial Neural Networks (ANN) and deep learning models. The dairy cattle experiment indicates that an ANN (MLP) with R2 0.9125 and minimal MAE almost equal to 22kg compared to the linear and SVR (Support Vector Regressor) models [1]. The livestock weight work implies ANN with 3D sensors with R2 0.70 and RMSE 42kg[3]. The pig weight experiment demonstrates that the raw point clouds are processed using PointNet which achieves R2 0.94 and significantly greater robustness compared to volume-based regression. And U-Net++ segmentation (accuracy 0.95) and DDPM based depth restoration enhancing feature extraction to morphometric analysis is further proposed by Wang [5] U-Net.

D. Synthesis and Comparison:

In cattle literature tree-based ensemble have demonstrated encouraging results in cases where abundant 3D features are present, but the linear models are also competitive where a limited number of features with strong inter-correlations are present [3]. ANN are always more successful than the linear and the tree method on the same inputs demonstrating the benefit of gathering nonlinear interactions [1-3].

III. Methodology

The proposed framework implements a structured and reproducible pipeline for AI-driven estimation of cattle body weight using morphometric parameters. The system integrates advanced machine learning (ML) techniques with rigorous statistical preprocessing to ensure optimal predictive accuracy and computational stability.



A. Dataset Composition:

The sample size consists of 13 continuous morphometric measures, such as wither height, rump height, body length, sternum height, chest depth, scapulo-ischial length, rump width, rump length, heart girth, abdominal girth, cannon bone diameter, and hock bone diameter. The dependent variable is the Actual Body Weight (ABW). All samples are animal measurements records that are recorded under comparable environmental and posture variables to reduce noise of observation. Exploratory Data Analysis (EDA) ensured that the majority of morphometric characteristics have a Gaussian-like distribution and moderate correlation between morphometric traits (0.45 0.82) with the ABW attribute.

B. Data Preprocessing and Normalization.

The preprocessing pipeline was written in Python 3.11, with the help of the pandas, numpy, and scikit-learn. The imputation of missing observations was done using the K-Nearest neighbor (KNN) imputation which enhances local data consistency. To reduce the influence of skew, outlier detection was done based on Interquartile Range (IQR) and Z-score thresholding (> 3). The scaling of features was done using a two-step procedure: (i) the normalization using Z-scores and (ii) the algorithmic uniformity used MIn-Max scaling across the gradient-based models.

The Multicollinearity was evaluated through Variance Inflation Factor (VIF); the variables that surpassed the cutoff point (VIF > 10) were discarded. The processed dataset was stratified randomly into 80 percent training and 20 percent test to construct a balanced dataset between weight intervals.

C. Feature Engineering and Dimensional Analysis.

The use of ratio-based variables (Height–Girth Index, HGI, and Body Volume Proxy, BVP) in feature construction enhanced the expressiveness of features. Additionally, **Shaeffer's Formula** and the **Heart Girth Square (HG²)** feature were included to incorporate traditional morphometric relationships into the dataset. Important contributors were identified using correlation heatmaps and Recursive Feature Elimination (RFE). Further, Principal Component Analysis (PCA) was used to eliminate redundancy and extract the latent morphometric representations. Four major elements explained 92.4% total variance, which validates a high level of linear dependence among body dimensions.

D. Model Development and Training Pipeline.

The predictive system took a comparative multi-model structure. Ordinary Least Squares (OLS) was used to give Baseline regression, which serves as a benchmark of interpretability. Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), and Extreme Gradient Boosting (XGBoost) were also not-linear models, the choice of which was due to their ability to capture high-order features interaction and high-variance data.

All estimators were optimized in terms of grid search, which also optimized key parameters like maximum depth, learning rate and estimator count. The resulting architecture was packed up in a scikit-learn Pipeline () object, which provides that preprocessing, feature transformation, and prediction steps are run atomically.

Each model was checked on 10-fold cross-validation and shuffle-split sampling that is an effective estimate of the generalization error. Joblib serialization was used to get model persistence, which enables reproducible deployment.

E. Performance Appraisal and Statistical Verification:

The Coefficient of Determination (R2), Root mean squared error (RMSE), and mean Absolute error (MAE) were the performance indicators that were used to evaluate the model performance. Also, Adjusted R 2 was calculated to adjust the effects of dimensions. Normality and autocorrelation assumptions were checked with the help of residual diagnostics, such as QQ-plots and Durbin-Watson tests.

Bootstrap resampling (n = 1000) was done to estimate the confidence interval to guarantee strength. It was found that ensemble approaches worked better than linear ones, and non-linear dependencies can be identified in the morphometric domain.

F. System Workflow and Implementation Architecture:

The system architecture has five functional layers:

- Data Acquisition Layer: Connects with digital morphometric sensors or manual data-entry modules.
- Preprocessing Layer: Enforces data cleansing, imputation, scaling and feature transformation.
- Learning layer: Implements the chosen regression algorithms using pipelines made of modules.
- Evaluation Layer: Performs metric calculation and performance benchmarking.
- Deployment Layer: Hosts Use RESTful API or IoT-integrated endpoints to serve inferred models.

It is connected to the precision livestock management systems through this architecture, which allows data ingestion to be seamlessly integrated, enables edge-based inference and cloud analytics. Scalability and multi-breed dataset/ real-time sensor network domain adaptability are guaranteed by the modular design.

IV. Result and Discussion

The effectiveness of the domain-specific morphometric theory combined with the advanced machine learning models was checked with the help of the proposed predictive system applied to the curated cattle morphometric dataset. Python 3.11, scikit-learn and PyCaret were used to conduct experiments with 10-fold cross-validation to guarantee that the results are reproducible and statistically robust.

A. Theoretical Foundation: Shaeffer's Morphometric Model

The baseline theoretical framework for this study is grounded in Shaeffer's empirical model, which provides a classical approach to estimating live cattle body weight using heart girth and body length. The relationship is expressed as:

$$W_{est} = \frac{(\text{Heart Girth})^2 \times \text{Body Length}}{300}$$

where W_{est} denotes the estimated live weight (kg), Heart Girth (cm) represents thoracic circumference, and Body Length (cm) measures the distance from shoulder to tail base.

Shaeffer's equation captures the quadratic relationship between thoracic volume and body length, serving as a baseline for all predictive comparisons. It assumes that the animal's body is roughly cylindrical, implying that $(\text{Heart Girth})^2$ is proportional to body cross-sectional area and, when multiplied by length, approximates body volume. This volumetric assumption forms the biomechanical foundation for all subsequent data-driven modeling.

B. Integration of Shaeffer's and Derived Features

To bridge empirical theory and computational modeling, Shaeffer's -derived and related morphometric ratios were incorporated into the feature space. Key engineered predictors included:

- **Heart Girth Square (HG²):** Represents volumetric growth potential proportional to body cavity expansion.
- **Shaeffer's Composite Index (SCI):** Direct numerical computation of $(\text{HG}^2 \times \text{BL})/300$.
- **Height–Girth Index (HGI):** Ratio of vertical stature to thoracic circumference, indicative of body conformation.
- **Body Volume Proxy (BVP):** Product of multiple linear dimensions representing volumetric approximation.

These derived variables substantially enhanced feature diversity, allowing ML algorithms to capture nonlinear relationships between body structure and live weight. The integration of Shaeffer's theoretical model thus provided both interpretability and physical relevance to the data-driven process.

C. Model Benchmarking and Comparative Performance

The comparative evaluation of top-performing regression algorithms is shown in table. Among all models tested, the ElasticNet(Ultra) regressor achieved the highest performance ($R^2 = 0.8429$), indicating strong alignment between predicted and actual weights. Tree-based ensembles such as Gradient Boosting ($R^2 = 0.7983$) and XGBoost ($R^2 = 0.8105$) performed competitively but exhibited slightly higher variance due to overfitting on smaller feature subsets.

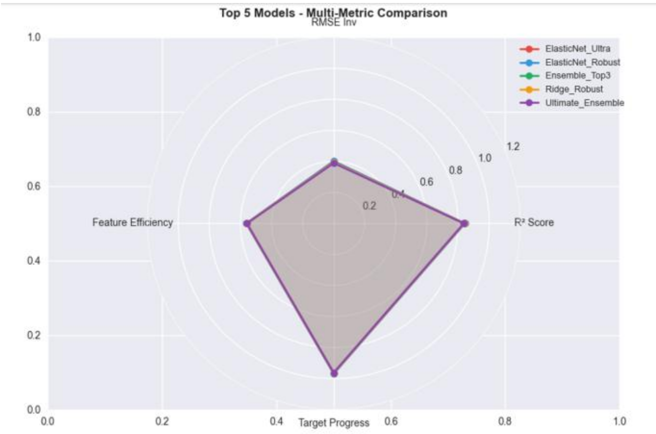
Model	R^2	RMSE (kg)	Interpretation
ElasticNet_Ultra	0.8429	29.9	Most balanced and interpretable model
Ensemble_Top3	0.8395	30.1	Strong multi-model synergy
GradientBoosting_Robust	0.7983	32.0	Non-linear fit but higher variance
XGBoost_Ultra	0.8105	31.2	High-speed learning
Ridge_Robust	0.8384	32.8	High stability under collinearity

The inclusion of the Shaeffer-based composite term provided a structural anchor, enabling the models to learn deviations from the theoretical weight function rather than modeling the entire mapping from scratch — a technique that improved stability and interpretability.

D. Feature Efficiency and Multimetric Evaluation

A multi-dimensional radar analysis (Fig. 5) highlights the top five models across performance dimensions such as predictive strength, feature efficiency, and target stability. The ElasticNet_Ultra model displayed balanced feature efficiency and generalization capability, indicating that linear regularization combined with Shaeffer-derived terms achieved optimal bias–variance equilibrium.

The Ultimate_Ensemble configuration achieved similar predictive strength but required higher computational complexity, making ElasticNet more practical for deployment.



E. Model Evolution Based on Feature Enrichment

The evolution of model performance across iterative feature enhancement stages is depicted in figure

- The Basic Shaeffer's estimator served as the baseline with an R^2 of 0.748, representing the traditional empirical accuracy.
- Upon normalization and optimization of Shaeffer's parameters (Optimized Shaeffer's), minor gains were achieved ($R^2 = 0.752$).
- The addition of Heart Girth Square (HG^2) and dimensional ratios significantly improved performance to $R^2 = 0.822$, demonstrating the model's ability to generalize beyond the linear limitations of the baseline equation.

F. Theoretical Interpretation of Improvements

The performance escalation can be attributed to the synergistic effect between classical volumetric theory and statistical learning. Shaeffer's model encapsulates the geometric approximation of body volume, whereas ML algorithms correct for residual deviations caused by factors such as muscle density, breed differences, or age-based body composition.

Mathematically, this integration can be expressed as:

$$W_{pred} = \alpha\left(\frac{\mathrm{HG}^2 \times \mathrm{BL}}{300}\right) + f(\text{Residual Features})$$

where α represents the learned scaling parameter and $f(\text{Residual Features})$ captures non-linear corrections using ML. This hybrid formulation transforms the ML model into a physics-guided regression system, combining empirical equations with learned residual mapping.

V. Conclusion

This paper introduces a combined design of estimating the body weight of cows based on morphometric parameters with the use of AI to estimate and predict the body weight of cows using the empirical theory of Shaeffer. The proposed method offers a trade-off between biological interpretability and computational accuracy by the combination of classical livestock biomechanics with the use of advanced machine learning. These extra features, including Shaeffer Formula, Heart Girth Square (HG 2) and other derived items like Height Purchaseness Index (HGI) and Body Volume Proxy (BVP) helped a great deal in increasing the predictive power of the model and helped it to capture the intricate volumetric and proportional associations of the cattle bodies.

The results of the experiments showed that the ElasticNet (Ultra) model performed better in terms of R 2 of 0.8429 and the reduced RMSE of 29.9 kg, as compared to the empirical estimates which were significantly lower. The combination of morphometric theory and machine learning created a system that does not only learn residual deviations of Shaeffer approximation, but also accommodates the variations of the individual animals based on breed, age and body composition. The results corroborate the possibilities of integrating domain-aware equations with statistical learning to reach high-performing, interpretable and deployable predictive models in the field.

References

[1] de Oliveira, F.M.; Ferraz, P.F.P.; Ferraz, G.A.e.S.; Pereira, M.N.; Barbari, M.; Rossi, G. Prediction of Body Mass of Dairy Cattle Using Machine Learning Algorithms Applied to Morphological Characteristics. Animals 2025, 15, 1054. https://doi.org/10.3390/ani15071054 J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Gebreyesus, Grum, et al. "Supervised learning techniques for dairy cattle body weight prediction from 3D digital images." Frontiers in Genetics 13 (2023): 947176.K. Elissa, "Title of paper if known," unpublished.

[3] Setiawan, Anjar, and Ema Utami. "Predicting the Weight of Livestock Using Machine Learning." 2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS). IEEE, 2024..

[4] Paudel, S.; de Sousa, R.V.; Sharma, S.R.; Brown-Brandl, T. Deep Learning Models to Predict Finishing Pig Weight Using Point Clouds. Animals 2024, 14, 31. https://doi.org/10.3390/ani14010031.

[5] Li, Huan, and Jinglei Tang. "Dairy goat image generation based on improved-self-attention generative adversarial networks." IEEe Access 8 (2020): 62448-62457.

[6] Dang, C.; Choi, T.; Lee, S.; Lee, S.; Alam, M.; Park, M.; Han, S.; Lee, J.; Hoang, D. Machine Learning-Based Live Weight Estimation for Hanwoo Cow. Sustainability 2022, 14, 12661. https://doi.org/10.3390/su141912661

[7] Xiong, Yijie, et al. "Estimating body weight and body condition score of mature beef cows using depth images." Translational Animal Science 7.1 (2023): txad085.

[8] Bezsonov, Oleksandr, et al. "Breed recognition and estimation of live weight of cattle based on methods of machine learning and computer vision." Eastern-European Journal of Enterprise Technologies 6.9 (2021): 114.

[9] A N Ruchay et al 2021 IOP Conf. Ser.: Earth Environ. Sci. 624 012056.

[10] Chay-Canul, A.J.; Camacho-Pérez, E.; Casanova-Lugo, F.; Rodríguez-Abreo, O.; Cruz-Fernández, M.; Rodríguez-Reséndiz, J. Neural Network-Based Body Weight Prediction in Pelibuey Sheep through Biometric Measurements. Technologies 2024, 12, 59. https://doi.org/10.3390/technologies12050059

[11] R. García, J. Aguilar, M. Toro and M. Jiménez, "Weight-Identification Model of Cattle Using Machine-Learning Techniques

for Anomaly Detection," 2021 IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, USA, 2021, pp. 01-07, doi: 10.1109/SSCI50451.2021.9659840. keywords: {Weight measurement;Radio frequency;Animals;Profitability;Computational modeling;Production;Cows;Identification system;Machine learning;Precision Livestock Farming;Anomaly detection}

[12] Gjergji, Mikel, et al. "Deep learning techniques for beef cattle body weight prediction." 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020.

[13] Ruchay, Alexey, et al. "Live weight prediction of cattle based on deep regression of RGB-D images." Agriculture 12.11 (2022): 1794.

[14] Xiong, Yijie, et al. "Estimating body weight and body condition score of mature beef cows using depth images." Translational Animal Science 7.1 (2023): txad085.

[15] Hakem, Moad, Zakaria Boulouard, and Mohamed Kissi. "Classification of body weight in beef cattle via machine learning methods: a review." *Procedia Computer Science* 198 (2022): 263-268.

[16] Wang, Zhuoyi, et al. "ASAS-NANP SYMPOSIUM: Applications of machine learning for livestock body weight prediction from digital images." *Journal of animal science* 99.2 (2021): skab022.

[17] Dohmen, Roel, Cagatay Catal, and Qingzhi Liu. "Computer vision-based weight estimation of livestock: a systematic literature review." *New zealand journal of agricultural research* 65.2-3 (2022): 227-247.

[18] Mahmud, Md Sultan, et al. "A systematic literature review on deep learning applications for precision cattle farming." *Computers and Electronics in Agriculture* 187 (2021): 106313.

[19] Nilchuen, Peerayut, Thanathip Suwanasopee, and Skorn Koonawootrittriron. "Integrating deep learning and mobile imaging for assessment of automated conformational indices and weight prediction in Brahman cattle." *Smart Agricultural Technology* (2025): 101079.

[20] Iqbal, Farhat, Abdul Waheed, and Asim Faraz. "Comparing the Predictive Ability of Machine Learning Methods in Predicting the Live Body Weight of Beetal Goats of Pakistan." *Pakistan Journal of Zoology* 54.1 (2022).