Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial intelligence-assisted prediction of body condition scores from radiographs of post-mortem free-roaming cats

Hyunji Jo
University of Florida
Ah-Young Kim
Animal and Plant Quarantine Agency
Donghee Lee
University of Florida
Ji-Su Baek
Animal and Plant Quarantine Agency
Kyunghyun Lee
Animal and Plant Quarantine Agency
BokKyung Ku
Animal and Plant Quarantine Agency
Adam W. Stern
University of Florida
Jong Hyuk Kim
jkim19@ufl.edu
University of Florida

Research Article

Keywords: artificial intelligence, body condition score, free-roaming cat, radiograph, veterinary forensic science

Posted Date: June 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9458448/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Body condition assessment is a critical component in evaluating general health and nutritional status, providing essential guidance for medical decision-making, welfare evaluation, and population-level monitoring. The body condition score (BCS) system has been widely adopted as a surrogate measure of body composition in animals. The aim of this study was to evaluate the feasibility of estimating BCS from radiographs of deceased free-roaming cats (FRCs) leveraging artificial intelligence (AI). We first established an automated image analysis pipeline to extract features from lateral whole-body radiographs of postmortem radiographs of FRCs (N = 257) using EfficientNetB3. The feature space was expanded tenfold through data augmentation (N = 2,570), and the dataset was divided into training (70%), validation (15%), and test (15%) sets at the source image level for each animal. We then compared multiple regression approaches to identify robust predictors of feline BCS and to provide a comprehensive evaluation of model performance.

Results

Kernel Ridge regression and Lasso were the top-performing models, exhibiting the highest explained variance on the test set (R^2 = 0.7199 and 0.7163, respectively) and the lowest prediction error (MAE = 0.6633 and 0.6645). Support Vector Regression and XGBoost showed intermediate performance, with reduced explanatory power (R^2 = 0.6498 and 0.6334) and higher error rates (MAE = 0.7468 and 0.7659). Random Forest exhibited further performance degradation (R^2 = 0.5639; MAE = 0.8202). The multi-layer perceptron model showed the weakest performance across all evaluated metrics (R^2 = 0.4584; MAE = 0.9251).

Conclusions

This study offers a technical validation of regression-based modeling for predicting postmortem BCS from radiographs of FRCs. Our comparative analysis revealed that regularized regression models, particularly Lasso, performed more effectively than more complex nonlinear models such as XGBoost and MLP. Future studies are needed to assess generalizability across diverse populations, imaging conditions, and prospective clinical or field settings.

Background

Body condition assessment is a critical component in evaluating general health and nutritional status, providing essential guidance for medical decision-making, welfare evaluation, and population-level monitoring [1, 2]. The gold standard for body composition measurements is dual-energy x-ray absorptiometry (DEXA), which offers an objective method for fat quantification and has been studied in

dogs and cats [3–5]. However, DEXA requires specialized equipment and resources, and anesthesia or sedation in live animals, limiting its practicality in general clinical settings. Body condition score (BCS) system has been widely adopted as a surrogate for body composition assessment since its introduction in cats three decades ago [1, 2, 6]. This method primarily estimates body fat mass through visual inspection and palpation of subcutaneous and abdominal fat deposits [1]. Its simplicity and accessibility across general practice and specialty care have contributed to its widespread clinical use. Previous studies have demonstrated significant positive correlations between DEXA-derived measurements and BCS, supporting BCS as a simple, reliable, and widely accepted tool for both clinical assessment and research applications, including studies of disease associations and health outcomes [2, 6–9]. Despite its long-standing and widespread use, BCS remains a subjective, semi-quantitative measure that is susceptible to inter-observer variability [2, 10]. Consequently, additional validation is required to clarify its performance and limitations in specific clinical and research settings.

The radiographic appearance of the vertebrae and surrounding soft tissues reflects body composition and fat distribution, suggesting that diagnostic imaging contains information relevant to BCS. For example, lateral abdominal radiographs measuring the falciform fat pad have been used to evaluate health conditions in conjunction with BCS in cats [10]. Another study examined the association between BCS and vertebral heart scale measurements on thoracic radiographs in dogs [11]. These studies highlight BCS as a practical and readily accessible clinical metric; however, the direct relationships between BCS and a wide range of health conditions remain incompletely characterized. In addition, most previous studies assessing the utility of BCS and its associations with other variables have been conducted in live animals. Free-roaming animals are exposed to numerous life-threatening hazards, including nutritional deficiencies, neglect (starvation), and vehicle-related trauma, and a high volume of forensic investigations to determine the cause of death are performed in cats [12–16]. As in living animals, body condition assessment is a critical initial step during forensic evaluations. Nevertheless, the potential utility of imaging data for BCS assessment in postmortem settings remain largely unexplored.

The aim of this study was to evaluate the feasibility of AI-driven, radiograph-based BCS estimation in carcasses of free-roaming cats (FRCs), with a primary focus on veterinary forensic applications. Here, we demonstrate that features extracted from whole-body radiographs of postmortem FRCs facilitate training of multiple regression models for predicting feline BCS.

Materials and Methods

Samples and BCS measurements

Autopsy records for 257 deceased FRCs submitted to the Animal Disease Diagnostic Division of the Animal and Plant Quarantine Agency (APQA) in the Republic of Korea between 2022 and 2025 were retrospectively reviewed. This study exclusively included feline carcasses submitted for forensic investigation; no live animals were involved, and no cats were euthanized for the purpose of this

research. Carcasses were examined immediately upon arrival or stored at 4°C prior to necropsy, with storage periods of less than two days. BCS was assessed using a 9-point scale through comprehensive body composition evaluation, including palpation of the rib cage and visual inspection of fat deposits throughout the body, with particular focus on the abdomen, lumbar region, and base of the tail during autopsy. Each carcass also underwent whole-body radiography, and the resulting digitized images were archived electronically. Institutional Animal Care and Use Committee (IACUC) approval was not required as all specimens were collected postmortem.

Overview of computational methodology

The schematic overview of this study, including image preprocessing, model development, and performance evaluation, is summarized in Fig. 1. The detailed methodology for each procedure is described in the section below.

Data preparation and preprocessing

The high-resolution lateral whole-body radiographs were originally acquired from carcass of FRCs, and images were standardized to select regions of interest. Images were manually cropped to include the caudal lumbar vertebrae, sacrum, proximal caudal (coccygeal) vertebrae, and pelvis. The lumbosacral-pelvic region was defined as the region of interest for radiographic analysis to ensure a standardized assessment of body condition. This region was prioritized because it offers a stable window to visualize epaxial muscle contours and subcutaneous fat deposits while remaining relatively unaffected by the superimposition of visceral organs and gas-induced artifacts. To improve computational efficiency, the resulting images were then processed through a preprocessing pipeline. Specifically, we implemented an image resizing procedure using the TensorFlow framework. The original radiograph images were converted into a fixed spatial resolution of 1024 · 1024 pixels. The resizing process employed bilinear interpolation to maintain the integrity of the pixel intensity values. Finally, images were encoded into a 3-channel RGB JPEG format to be compatible with the input requirements of pre-trained convolutional neural networks originally trained on the ImageNet database [17, 18].

Feature extraction and data augmentation

An offline data augmentation strategy was employed to artificially expand the training distribution prior to model training. Using a sequential preprocessing layer, we applied random transformations including rotations (up to 10%), zooms (10%), and width/height shifts (10%) to each input image. For each original radiograph image, ten augmented versions (N = 10) were pre-generated, ensuring that the regressor could learn from a diverse range of spatial variations (Fig. 2). Second, for deep feature extraction, we utilized the EfficientNetB3 architecture pre-trained on the ImageNet dataset. To accommodate the high-resolution 1024 · 1024 inputs and ensure architectural integrity, we adopted a two-step weight loading procedure. The model structure was first initialized with the specific input dimensions, after which the

pre-trained weights were manually mapped to the corresponding layers. The final classification layer was replaced with a Global Average Pooling (GAP) layer, transforming the high-dimensional feature maps into a 1,536-dimensional feature vector. This process enabled the extraction of high-level semantic features that may represent feline body composition in radiographic images while maintaining numerical stability for subsequent regression analysis.

The extraction process was executed through an automated iterative pipeline. For each input image, we implemented a sequential augmentation-and-extraction loop. Each standardized image was subjected to the predefined augmentation suite ten times, followed by model-specific intensity scaling using the EfficientNet preprocessing function. This function ensures that input pixel values are normalized to a range compatible with the model's pre-trained weights (typically scaled to [0, 1] or standardized to Z-scores), which is critical for maintaining the numerical stability of the extracted deep features.

The preprocessed image batches were then fed into the EfficientNetB3 backbone. By applying GAP at the final layer, the spatial feature maps were collapsed into a one-dimensional feature vector of length 1,536 for each augmented instance. These high-dimensional vectors were flattened and aggregated into a comprehensive feature matrix. This exhaustive sampling approach resulted in an expanded feature space, providing a richer data representation for the subsequent regression analysis. The finalized feature and label arrays were serialized as NumPy binary files to ensure reproducible input for the model optimization stage.

Dataset partitioning

A structured dataset partitioning strategy was applied based on the original, non-augmented image indices. Although the feature space was expanded tenfold through data augmentation (N = 2,570), the split was strictly controlled at the source image level for each animal (N = 257). The dataset was divided into three subsets: training (70%, N = 179 original images), validation (15%, N = 39 images), and test (15%, N = 39 images). By partitioning the data based on the original indices prior to mapping them to their respective augmented instances, we ensured that no variations of the same original radiograph appeared across different sets. Following the index-based split, the corresponding augmented feature vectors were grouped to form the final matrices. This resulted in a training pool of 1,790 instances and validation and test sets of 390 instances each. These partitioned datasets were then used for subsequent hyperparameter optimization and regression analysis to ensure consistent evaluation across all models.

Regression models

To identify a robust estimator for feline BCS prediction, we evaluated a diverse suite of regression architectures, including linear, kernel-based, ensemble, and neural network approaches. Specifically, we implemented Lasso (for built-in feature selection) [19]; ensemble methods, including Random Forest [20] and XGBoost [21], to capture non-linear interactions in the high-dimensional feature space; kernel-based

methods, including Support Vector Regression (SVR) [22] and Kernel Ridge Regression [23], to map features into higher-dimensional spaces for improved separability; and a multi-layer perceptron (MLP) [24] to model complex non-linear relationship. Using this approach, we compared BCS prediction performance across distinct algorithmic paradigms.

Hyperparameter optimization and model selection

For each regression architecture, we implemented a grid search with k-fold cross-validation (k = 10). This approach involved defining a comprehensive search space for key parameters. To support the effectiveness of the optimization, we conducted a performance-driven selection process. The hyperparameter-tuned model was compared against the baseline model (with default parameters). The optimized configuration was adopted as the final model only if it achieved a lower Mean Squared Error (MSE) than the baseline performance, thereby confirming that the tuning process yielded a statistically superior representation.

Performance evaluation metrics

To evaluate the predictive accuracy and robustness of the regression models, we employed five standard statistical metrics: Mean Absolute Error (MAE), MSE, Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE).

MAE: Measures the average magnitude of the errors in a set of predictions, without considering their direction.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y_i}|$$

MSE: Quantifies the average of the squares of the errors, placing higher penalties on larger deviations.

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \widehat{y_i})^2$$

RMSE: Provides the error magnitude in the same units as the target variable (BCS), facilitating intuitive interpretation.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}$$

R^2

Represents the proportion of variance in the BCS scores that is predictable from the deep features, indicating the overall goodness-of-fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

MAPE: Expresses the prediction accuracy as a percentage, allowing for an assessment of the relative error scale.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right|$$

Consistent evaluation criteria were applied to all architectures, including Lasso, XGBoost, Random Forest, SVR, Kernel Ridge, and MLP, to ensure that the comparative analysis of feature extractors and regressors was statistically valid.

Computational environment

All experiments were conducted in a Python 3.10.13 environment. The machine learning based regression models including Lasso, Random Forest, SVR, and Kernel Ridge were implemented using scikit-learn (version 1.5.0). For the gradient boosting analysis, the XGBoost (version 3.0.5) library was utilized. Both the feature extraction via EfficientNetB3 and the deep learning-based MLP were developed and executed using the TensorFlow (version 2.20.0) framework. Data preprocessing and feature matrix management were performed using Numpy (version 2.2.6) and Pandas (version 2.2.2). All computational tasks were executed on the University of Florida (UF) HiPerGator high performance computing cluster using CPU based nodes.

Statistical analysis

Statistical analyses were performed to characterize the distribution of the BCS data. Descriptive statistics, including mean, standard deviation, and frequency distributions, were computed to summarize the dataset using GraphPad Prism software (Version 11.0.0).

Results

Descriptive statistics of datasets

We first analyzed the distribution of BCS values prior to model selection and development. BCS was measured on a 9-point scale (ranging from 1 to 9), as represented in Fig. 3A. Although the QQ plot suggested an approximately normal distribution, relatively few samples were observed at the extreme scores (i.e., near 1 or 9) (Figs. 3B, C). Most cases were concentrated in the mid-range (mean = 4.658; standard deviation = 1.581), with 161 samples (62.6%) scoring between 4 and 6, and 88.7% of scores falling between 3 and 7.

Hyperparameter optimization for model development

Next, we optimized hyperparameters during the model development for predicting feline BCS. Table 1 summarizes the comparative performance of six regression models before and after the grid search-based hyperparameter optimization. Among the evaluated models, XGBoost achieved the lowest error rate with a MSE of 1.0964, utilizing a learning rate of 0.05 and a maximum depth of 3. This indicates that the gradient boosting approach effectively captured the non-linear relationships within the deep features. The Lasso regression exhibited the most significant performance improvement through optimization: the baseline MSE of 2.6147 was reduced by approximately 55% to a final MSE of 1.1825 (α = 0.1). This substantial gain suggests that the high-dimensional feature space extracted from EfficientNet contained redundant information, which was effectively managed by Lasso's L1 regularization penalty. Kernel Ridge also showed a marked improvement (from MSE 1.5750 to 1.2012), while Random Forest and SVR demonstrated relatively stable performance with marginal gains over their baselines. In contrast, the deep learning-based regressor showed a comparatively higher error rate. For the MLP model, a baseline MSE is not reported as the architecture was directly optimized through iterative design (incorporating batch normalization and dropout) rather than adjusting from a standardized default configuration. Despite the structural optimization (MSE 1.4685), the simpler linear and ensemble models outperformed the MLP in this specific task, likely due to the limited sample size relative to the model complexity. For Lasso and Kernel Ridge, despite the ordinal and non-normally distributed nature of the BCS data, hyperparameter tuning substantially improved MSE, suggesting that regularization mitigated potential distributional issues.

Table 1
Comparison of MSE between baseline and hyperparameter-tuned models.

Regression models	Baseline MSE	Best MSE	Best hyperparameters
Lasso	2.6147	1.1825	α = 0.1
XGBoost	1.3854	1.0964	lr = 0.05, depth = 3, n_{est}=400, sub = 0.8
Random Forest	1.2632	1.2503	depth = 20, n_{est}=200, leaf = 2, split = 2
Kernel Ridge	1.5750	1.2012	α = 0.1, γ = 0.0001
SVR	1.3131	1.2895	C = 10, γ = auto, kernel = rbf
MLP	n/a	1.4685	Arch:32-16-1, λ=0.05, p = 0.5, lr = 0.0005

n/a = not applicable; lr = learning rate; depth = max_depth; n_{est} = n_estimators; sub = subsample; γ = kernel coefficient; C = regularization parameter; kernel = kernel type; p = dropout probability; α = alpha (regularization strength); Arch = network architecture; λ = lambda (L2 regularization)

Taken together, our work shows that hyperparameter tuning effectively reduced the prediction error across most architectures, facilitating a performance-driven selection process.

Comparative performance of BCS prediction models

We then evaluated the predictive performance of six regression models with optimized hyperparameters using an independent hold-out test set. The comparative analysis revealed performance differences among regression families (Table 2; Fig. 4). Overall, models incorporating explicit regularization achieved superior generalization performance compared with ensemble-based and deep learning approaches. In particular, Kernel Ridge regression and Lasso emerged as the top-performing models, exhibiting the highest explained variance on the test set (R^2 = 0.7199 and 0.7163, respectively) and the lowest prediction error (MAE = 0.6633 and 0.6645). SVR and XGBoost showed intermediate performance, with reduced explanatory power (R^2 = 0.6498 and 0.6334) and higher error rates (MAE = 0.7468 and 0.7659). Random Forest exhibited further performance degradation (R^2 = 0.5639; MAE = 0.8202). The MLP model showed the weakest performance across all evaluated metrics (R^2 = 0.4584; MAE = 0.9251), despite its greater model complexity. To further characterize model behavior beyond aggregate error metrics, we examined the relationship between actual and predicted BCS values for the two best-performing models (Fig. 5). Our data showed that both Lasso and Kernel Ridge predictions exhibited strong positive correlations with observed BCS, with data points closely aligned along the identity line, indicating good overall calibration and low systematic bias. Markedly, the Lasso model demonstrated a particularly stable linear relationship across the full BCS range (2–9), with fewer large deviations at extreme values.

Table 2
BCS prediction performance of regression models on a hold-out test set

Models	R^2	MAE	MSE	RMSE	MAPE
Kernel Ridge	0.7199	0.6633	0.7593	0.8714	13.98%
Lasso	0.7163	0.6645	0.7693	0.8771	13.95%
SVR	0.6498	0.7468	0.9496	0.9745	16.35%
XGBoost	0.6334	0.7659	0.9940	0.9970	16.82%
Random Forest	0.5639	0.8202	1.1825	1.0874	18.06%
MLP	0.4584	0.9251	1.4685	1.2118	18.29%

Collectively, these results suggest that regularized regression models, particularly Lasso, generalized better to unseen data than more complex models, achieving accurate and stable BCS predictions across the full scoring range under the current data constraints.

Discussion

This study demonstrates that radiographic features extracted from feline lumbosacral radiographs can be used to predict BCS using regression models. Through comparative analyses, our data show that regularized linear and kernel-based methods outperform more complex non-linear and deep learning regressors on our dataset. This work provides a feasibility evaluation and technical validation of regression approaches in the context of feline medicine and veterinary forensic science.

One of our main findings is that simpler, regularized models, specifically Lasso and Kernel Ridge, achieve superior performance compared to more complex non-linear models such as XGBoost and MLP. This is likely attributable to effective capacity control regularization, which is particularly important when modeling 1,536-dimensional feature vectors derived from a relatively limited sample size. This observation is consistent with the well-established high-dimensional, low-sample-size (HDLSS) challenge in medical image analysis, where parsimonious models often outperform high-capacity approaches by reducing overfitting to redundant or noisy features [25–29]. In addition, the feature representations extracted using the EfficientNetB3 backbone may contribute to this behavior, as such deep convolutional networks can produce high-dimensional embeddings that contain both informative structure and redundant noise [18, 30, 31]. In this regard, Lasso regression provides effective complexity control through L1 regularization, which induces sparsity by shrinking irrelevant coefficients to zero and thereby performing embedded feature selection [19, 32, 33]. Kernel Ridge Regression, in contrast, imposes smoothness constraints that stabilize predictions and help control variance while preserving relevant structure in the data [19, 34]. By comparison, ensemble methods and MLP architectures possess higher representational capacity, which can increase susceptibility to overfitting in small-sample, high-dimensional settings despite hyperparameter tuning. This is reflected in their lower R^2 and

higher prediction error on hold-out data [35–37]. Although the MLP model incorporated L2 regularization (λ = 0.05) and dropout (p = 0.5), L2 regularization primarily shrinks coefficient magnitudes rather than enforcing sparsity, which may limit its ability to fully suppress cumulative noise in high-dimensional feature spaces. Overall, these results suggest that when sample size is limited relative to feature dimensionality, performance is improved by prioritizing strong regularization and capacity control rather than increasing model complexity. This supports the practical advantage of simpler regression frameworks for achieving robust generalization in veterinary imaging applications [38].

Our results provide insights with potential clinical and research relevance. First, this study indicates that radiographs contain extractable and relevant information related to feline body condition, supporting the feasibility of automated BCS estimation in situations where physical scoring is unavailable or where only imaging data are available, such as in retrospective surveillance and population-level screening. In veterinary forensic science, shelter medicine, and population health programs, such tools could help identify underweight or overweight cohorts within archival radiograph datasets, support prioritization of nutritional or medical interventions, and provide objective and reproducible estimates to complement routine intake assessments. While postmortem imaging (such as radiographs) are required to be performed in all cases of gunshot injuries and charred remains in animals ≤ 70 kg [39], as radiographic analysis becomes more routine in veterinary forensic investigations, automated BCS scoring using radiographs can become commonplace and would be additional data that can be obtained before performing the postmortem examination. Nonetheless, since the current model was developed using deceased FRCs, direct clinical translation will require prospective validation in live populations and calibration to institution-specific case distributions and imaging protocols. Ethical and operational considerations should be addressed prior to deployment in decision-support workflows.

Furthermore, the observed advantage of regularized models offers a practical modeling guideline for similar HDLSS problems in veterinary imaging, suggesting that simpler, well-regularized regressors or kernel-based methods may be preferable when feature dimensionality substantially exceeds available sample size. The use of regularized models removes interobserver variability as it maintains consistency between analyses and relies on objective data rather than subjective criterion used during palpation and visual inspection of the body.

Several limitations should be considered when interpreting these findings. The dataset consists of deceased FRCs, which may differ from live clinical populations in body composition, positioning, and soft-tissue characteristics. The relatively small sample size limits the training of high-capacity models and constrains generalizability. In this dataset, 88.7% of BCS fell between 3 and 7, with relatively few animals represented at the extremes of 1 and 9. In addition, the use of single-view lateral radiographs may not capture the full spectrum of anatomical features relevant to clinical BCS assessment, and postmortem changes may introduce additional variability in image-derived features.

Conclusions and future directions

In conclusion, we show that an automated image analysis pipeline can be established and that AI-assisted BCS prediction from radiographic images is achievable using regression approaches. This work is presented within the context of veterinary forensic science, with a focus on FRCs, and offers a conceptual and technical framework for the further development of AI-driven applications in veterinary medicine.

Future studies should focus on validation using larger, prospectively collected datasets that include live clinical cases with standardized and comprehensive BCS scoring. In cases of feline starvation, cats are typically expected to present at the lowest extreme (BCS 1/9), which is of particular forensic significance. Future studies should aim to include a greater number of cases at this extreme to better evaluate its diagnostic and investigative value. Expanding the framework to incorporate multi-view radiography or complementary imaging modalities may further improve robustness and predictive performance. Integrating image-derived features with basic clinical metadata, including age, sex, and body weight, and health-related information may also enhance model accuracy while preserving interpretability and clinical usability. Finally, external validation across institutions, equipment, facilities, and imaging protocols will be essential to confirm model robustness across diverse settings.

Declarations

Acknowledgments

We acknowledge the UF HiPerGator Research Computing for computational resources and technical support.

Author contributions statement

AYK, KL, BKK, and JHK conceptualized and designed the study. HJ, AYK, DHL, and JSB prepared samples, collection and processing of datasets, and conducted the research. KL, BKK, AWS, and JHK interpreted and analyzed the data. HJ and JHK wrote original manuscript. All authors contributed to writing and revising the manuscript, reviewed the final manuscript, approved its submission, and agree to be accountable for all aspects of the work.

Disclosure statement

The authors declare that this research was conducted without any commercial or financial relationships that could be interpreted as a potential conflict of interest. Generative AI tools, including GPT-5.2 and Gemini 3, were used solely for language editing and proofreading, including correction of typographical and grammatical errors. All sentences and textual content were strictly and thoroughly reviewed and additionally edited by the authors.

Funding

This work was partly supported by the UF Foundation “A Cat Has No Name” fund and the UF AI Initiative. This work was also supported by a grant (I-1543069-2024-26-01) from the Animal and Plant Quarantine Agency (APQA), Ministry of Agriculture, Food and Rural Affairs, Republic of Korea.

Consent to Publish declaration

Not applicable.

Availability of data and materials

The original contributions presented in this study are included in the manuscript. Further inquiries can be directed to the corresponding author. The data supporting the conclusions of this article may be made available by the corresponding author upon request, subject to institutional guidelines and a data use agreement.

References

1. Laflamme D. Development and validation of a body condition score system for cats: a clinical tool. Feline Pract. 1997;25(5/6):13–8.
2. German AJ, Holden SL, Moxham GL, Holmes KL, Hackett RM, Rawlings JM. A simple, reliable tool for owners to assess the body condition of their dog or cat. J Nutr. 2006;136(7 Suppl):S2031–3.
3. Lauten SD, Cox NR, Baker GH, Painter DJ, Morrison NE, Baker HJ. Body composition of growing and adult cats as measured by use of dual energy X-ray absorptiometry. Comp Med. 2000;50(2):175–83.
4. Speakman JR, Booles D, Butterwick R. Validation of dual energy X-ray absorptiometry (DXA) by comparison with chemical analysis of dogs and cats. Int J Obes Relat Metab Disord. 2001;25(3):439–47.
5. Hammami M, Koo WW, Hockman EM. Body composition of neonates from fan beam dual energy X-ray absorptiometry measurement. JPEN J Parenter Enter Nutr. 2003;27(6):423–6.
6. Bjornvad CR, Nielsen DH, Armstrong PJ, McEvoy F, Hoelmkjaer KM, Jensen KS, Pedersen GF, Kristensen AT. Evaluation of a nine-point body condition scoring system in physically inactive pet cats. Am J Vet Res. 2011;72(4):433–7.
7. Peterson ME, Castellano CA, Rishniw M. Evaluation of Body Weight, Body Condition, and Muscle Condition in Cats with Hyperthyroidism. J Vet Intern Med. 2016;30(6):1780–9.
8. Teng KT, McGreevy PD, Toribio JALM, Raubenheimer D, Kendall K, Dhand NK. Associations of body condition score with health conditions related to overweight and obesity in cats. J Small Anim Pract 2018.
9. Pye CR, Dowgray NJ, Eyre K, Pinchbeck G, Biourge V, Moniot D, Comerford E, German AJ. Longitudinal changes in bodyweight, body condition, and muscle condition in ageing pet cats: findings from the Cat Prospective Ageing and Welfare Study. Front Vet Sci. 2025;12:1654002.

10. Scott KC, Levy JK, Gorman SP, Newell SM. Body condition of feral cats and the effect of neutering. J Appl Anim Welf Sci. 2002;5(3):203–13.
11. Bruining-Staal K, Tiemessen I, Vernooij JCM, Beijerink N. Effect of body condition score on the interobserver variability of vertebral heart scale assessment in dogs. Vet Radiol Ultrasound. 2024;65(6):694–701.
12. Bruce SJ, Zito S, Gates MC, Aguilar G, Walker JK, Goldwater N, Dale A. Predation and Risk Behaviors of Free-Roaming Owned Cats in Auckland, New Zealand via the Use of Animal-Borne Cameras. Front Vet Sci. 2019;6:205.
13. de Siqueira A, Cassiano FC, de Albuquerque Landi MF, Marlet EF, Maiorka PC. Non-accidental injuries found in necropsies of domestic cats: a review of 191 cases. J Feline Med Surg. 2012;14(10):723–8.
14. Almeida DC, Torres SMF, Wuenschmann A. Retrospective analysis of necropsy reports suggestive of abuse in dogs and cats. J Am Vet Med Assoc. 2018;252(4):433–9.
15. Thuesen IS, Agerholm JS, Mejer H, Nielsen SS, Sandøe P. How Serious Are Health-Related Welfare Problems in Unowned Unsocialised Domestic Cats? A Study from Denmark Based on 598 Necropsies. Anim (Basel) 2022, 12(5).
16. Nielsen SS, Thuesen IS, Mejer H, Agerholm JS, Nielsen ST, Jokelainen P, Thamsborg SM, Sandøe P. Assessing welfare risks in unowned unsocialised domestic cats in Denmark based on associations with low body condition score. Acta Vet Scand. 2023;65(1):1.
17. Deng J, Dong W, Socher R, Li LJ, Kai L, Li F-F. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition: 20–25 June 2009*; 2009: 248–255.
18. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *ArXiv* 2019, abs/1905.11946.
19. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. J Roy Stat Soc: Ser B (Methodol). 2018;58(1):267–88.
20. Breiman L. Random Forests. Mach Learn. 2001;45(1):5–32.
21. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* San Francisco, California, USA: Association for Computing Machinery; 2016: 785–794.
22. Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support vector regression machines. Adv Neural Inf Process Syst 1997:155–61.
23. Saunders C, Gammerman A, Vovk V. Ridge Regression Learning Algorithm in Dual Variables. In: *International Conference on Machine Learning*: 1998; 1998.
24. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986;323(6088):533–6.
25. Altman N, Krzywinski M. The curse(s) of dimensionality. Nat Methods. 2018;15(6):399–400.

26. Papanikolaou N, Matos C, Koh DM. How to develop a meaningful radiomic signature for clinical use in oncologic patients. Cancer Imaging. 2020;20(1):33.
27. Chicco D, Shiradkar R. Ten quick tips for computational analysis of medical images. PLoS Comput Biol. 2023;19(1):e1010778.
28. Berisha V, Krantsevich C, Hahn PR, Hahn S, Dasarathy G, Turaga P, Liss J. Digital medicine and the curse of dimensionality. NPJ Digit Med. 2021;4(1):153.
29. Johnstone IM, Titterington DM. Statistical challenges of high-dimensional data. *Philos Trans A Math Phys Eng Sci* 2009, 367(1906):4237–4253.
30. Raghu M, Zhang C, Kleinberg JM, Bengio S. Transfusion: Understanding Transfer Learning for Medical Imaging. In: *Neural Information Processing Systems: 2019*; 2019.
31. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24–9.
32. Tibshirani R, Saunders MA, Rosset S, Zhu J, Knight K. Sparsity and smoothness via the fused lasso. J Royal Stat Society: Ser B (Statistical Methodology) 2005, 67.
33. Fira M, Goras L, Costin H-N. Evaluating Sparse Feature Selection Methods: A Theoretical and Empirical Perspective. Appl Sci. 2025;15(7):3752.
34. Scholkopf B, Smola A. Learning with Kernels: support vector machines, regularization, optimization, and beyond. In: *Adaptive computation and machine learning series: 2001*; 2001.
35. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning (still) requires rethinking generalization. Commun ACM. 2021;64(3):107–15.
36. Dietterich TG. Ensemble Methods in Machine Learning. Multiple Classifier Systems: 2000// 2000; Berlin. Heidelberg: Springer Berlin Heidelberg; 2000. pp. 1–15.
37. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15(1):1929–58.
38. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. J Royal Stat Soc Ser B: Stat Methodol. 2005;67(2):301–20.
39. Standard for Veterinary Forensic Postmortem Examination. 1st edn: American National Standards Institute/Acadeny Standards Board; 2024.

Figures

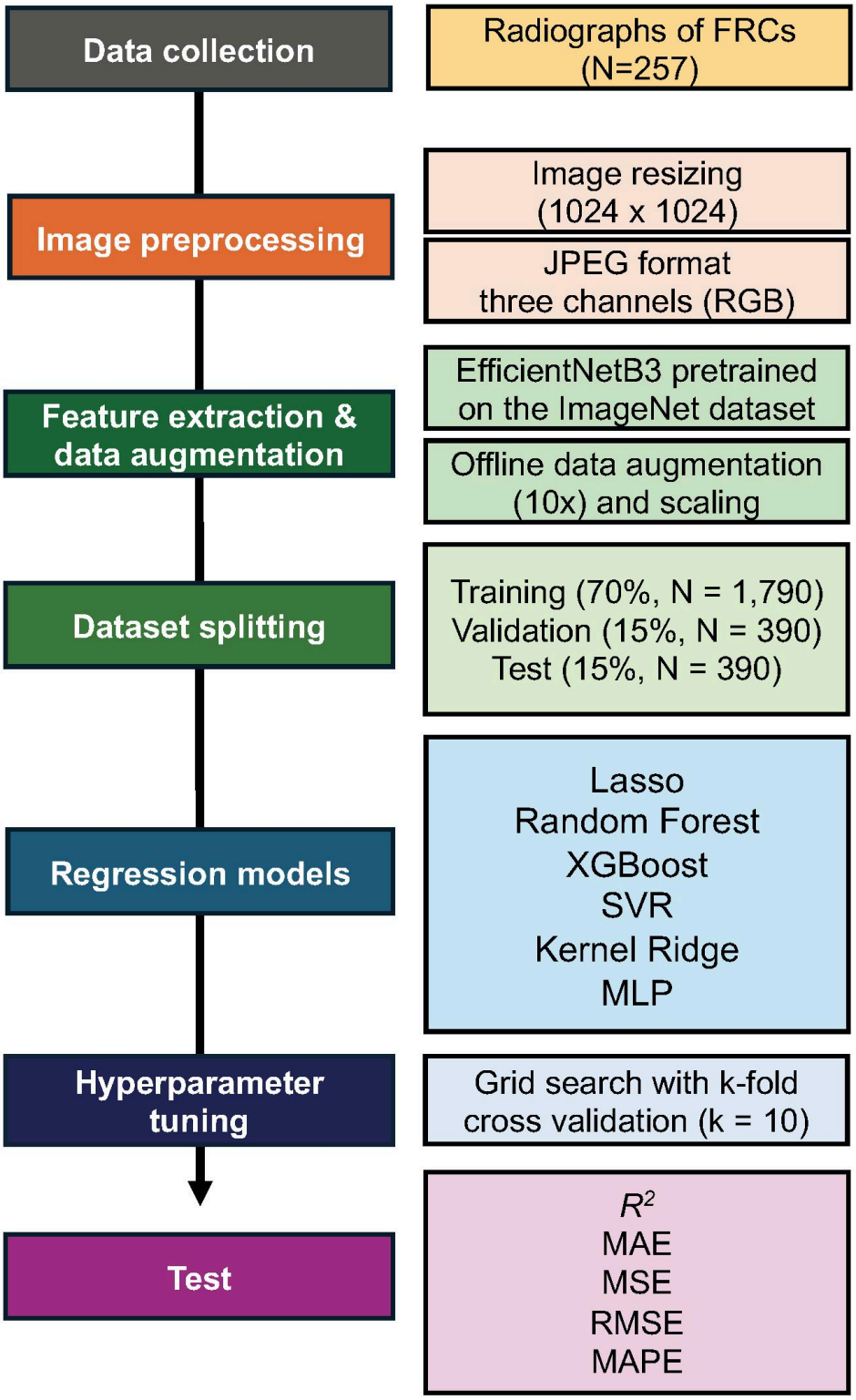


Figure 1

Schematic overview of this study, including image preprocessing, model development, and performance evaluation.

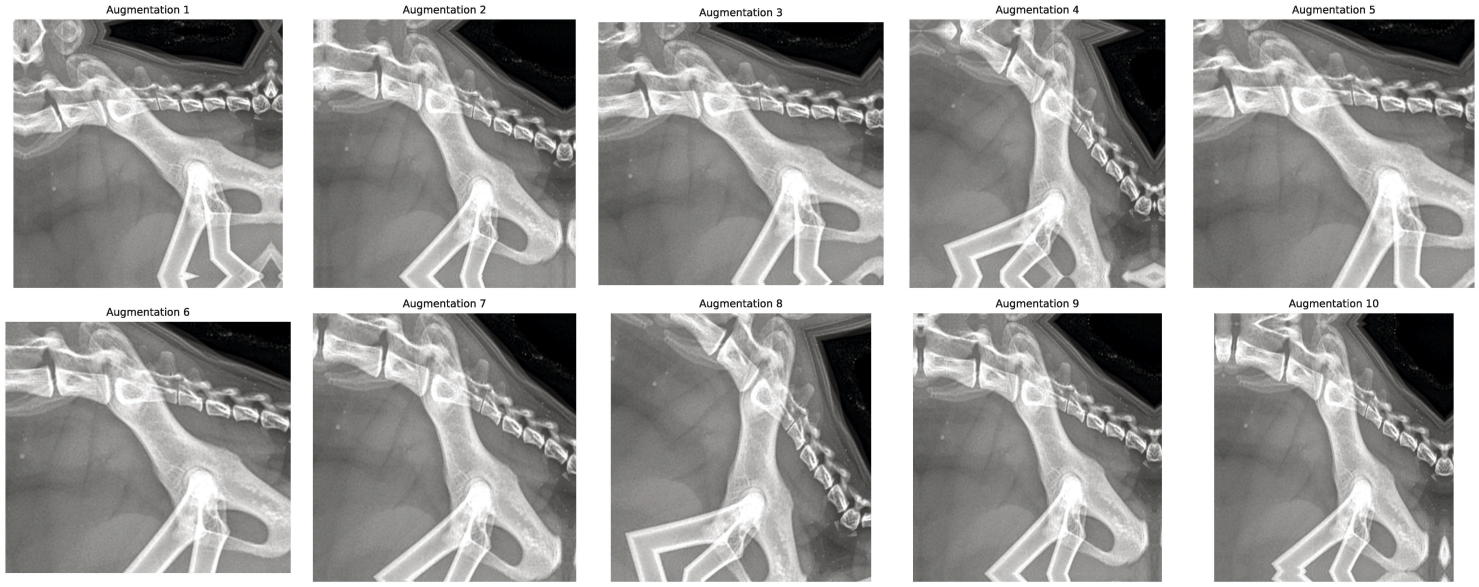


Figure 2

Representative radiographic image dataset illustrating random data augmentation transformations, including rotations, zooming, and width and height shifts, resulting in ten augmented versions.

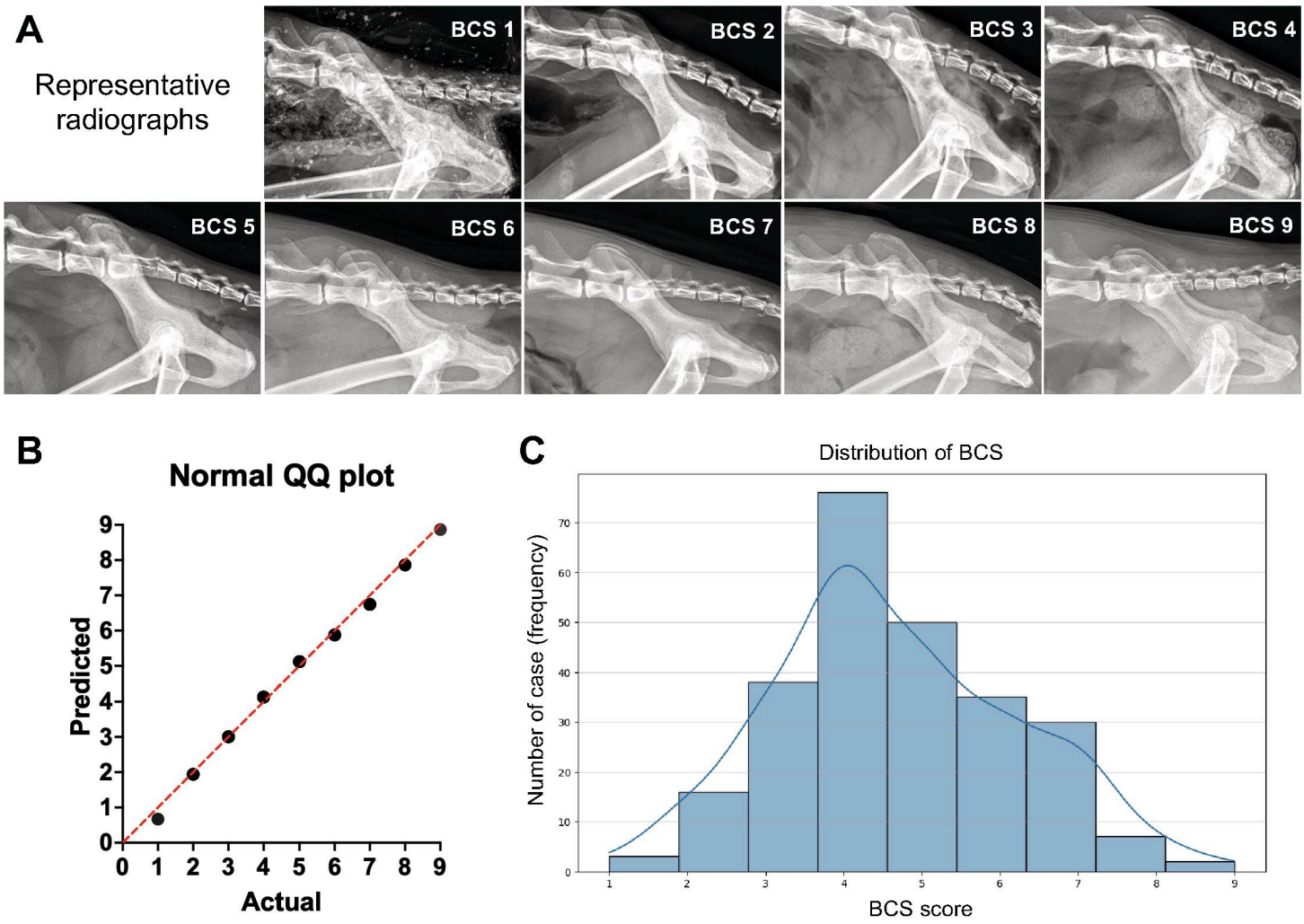


Figure 3

(A) Representative radiographs of autopsy cases of FRCs labeled with different BCS values are shown. (**B**) QQ plot and (**C**) histogram show the distribution of the data.

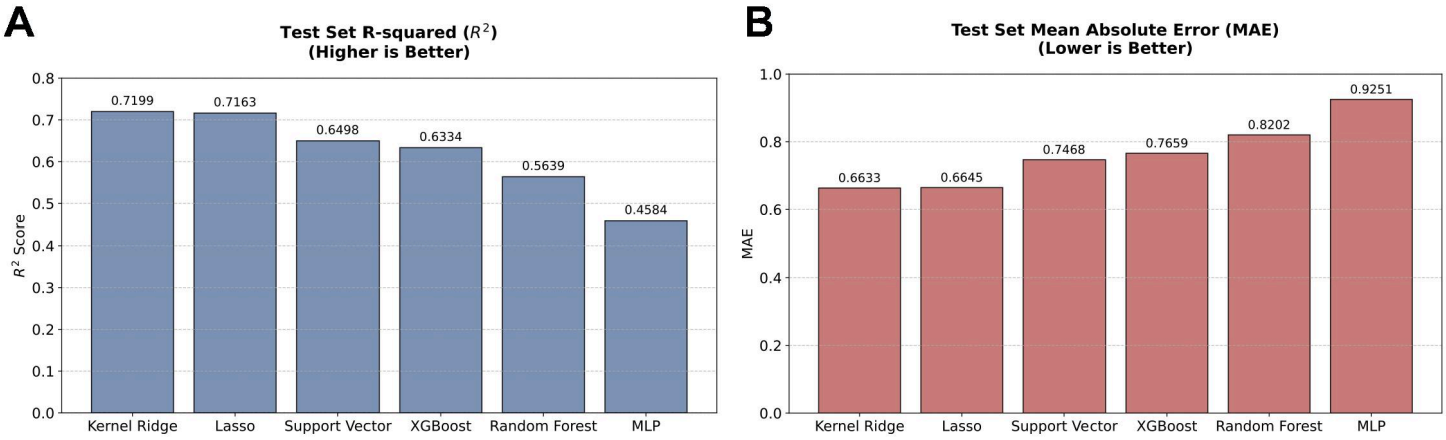


Figure 4

Comparative performance of six regression models on the test set. Bar plots illustrate differences in predictive performance across models using R^2 (**A**) and MAE (**B**).

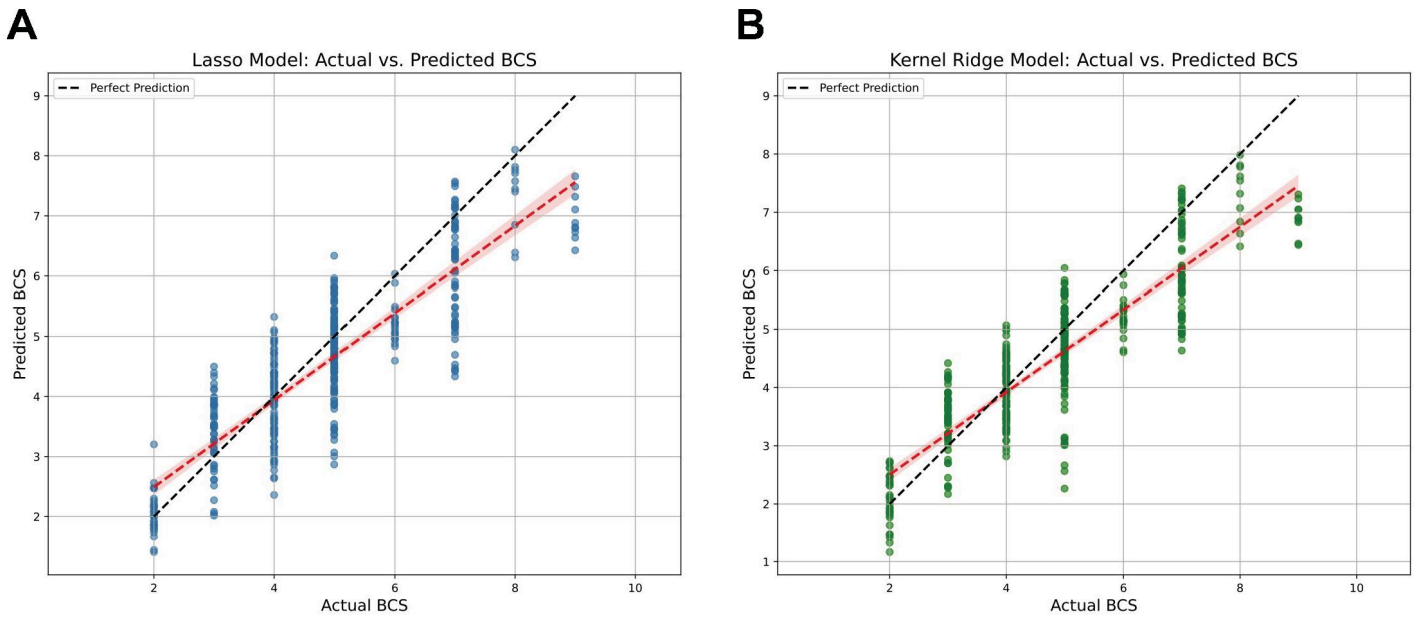


Figure 5

Scatter plots of actual versus predicted BCS scores for the top two models: (**A**) Lasso and (**B**) Kernel Ridge. The proximity of the data points to the perfect prediction line (black dashed) and the alignment of the trend line (red) indicate high predictive accuracy and low bias.