Weight Trajectory Prediction in Precision Livestock Farming Using Machine Learning: A Comparative Approach

Moad Hakem, Zakaria Boulouard, Mohamed Kissi
Faculty of Sciences and Technologies, Department of Computer Science, Hassan II University, Mohammedia, Morocco

***Abstract*—Accurate livestock body weight prediction is a key component of precision livestock farming, as it supports herd monitoring, production management, and planning in response to the increasing global demand for meat. Existing approaches for weight prediction include age-based regression models, growth trajectory modelling, average daily gain estimation, and methods relying on morphometric measurements or image-derived features. However, many of these approaches require frequent measurements or specialized data acquisition systems, which are often costly and difficult to deploy under practical farming conditions. This study presents a comparative evaluation of data-driven models for livestock body weight trajectory prediction under low-measurement conditions. A matrix factorization approach and four ensemble-based machine learning methods, namely XGBoost, LightGBM, CatBoost, and ExtraTrees, were evaluated using a dataset of Holstein cows. Model performance was assessed using standard regression metrics, including root mean squared error, mean absolute error, and mean absolute percentage error, with five-fold cross-validation employed to ensure robustness. The results show that ensemble learning methods consistently outperform matrix factorization techniques when only a limited number of weight measurements per animal are available. More specifically, XGBoost achieves the best predictive performance when only one historical measurement per animal is available, whereas ExtraTrees provides the most accurate predictions when two or three historical measurements are available. These findings demonstrate that accurate and cost-effective livestock weight prediction can be achieved from sparse routine body weight records, without relying on dense longitudinal sampling, image-based systems, or extensive morphometric measurements, thereby supporting the practical deployment of predictive tools in precision livestock farming systems.**

***Keywords*—Machine learning; data science; precision livestock farming; weight trajectory prediction; ensemble learning**

I. Introduction

In livestock species, body weight is a key biological variable that characterizes growth trajectories and reflects the physiological processes governing animal development throughout the life cycle [1]. From a production perspective, body weight gain represents a fundamental performance trait in cattle, as it is closely associated with feed intake and growth efficiency and is widely used to assess animal performance in beef systems [2]. As a result, longitudinal body weight records are widely exploited in livestock management to support feeding decisions, assess growth performance, and identify optimal slaughter timing. Beyond productivity considerations, longitudinal variations in body weight constitute a valuable source of information for animal health monitoring, as atypical weight patterns may be associated with metabolic imbalances or pathological conditions [3]. In addition, body weight records are commonly exploited in animal evaluation studies, where growth-related information derived from weight trajectories contributes to the assessment of production-related traits across individuals [4].

Despite the central role of body weight in livestock management, acquiring systematic and frequent live-weight measurements remains challenging under practical farming conditions. In most production systems, animals are weighed at only a limited number of time points throughout their growth period, resulting in longitudinal datasets that are often sparse and irregular. Such data limitations hinder the accurate characterization and prediction of individual growth trajectories, as reported in previous studies on cattle growth modelling [5], [6], [7]. Traditional live-weight measurement in cattle production relies primarily on either manual handling procedures or industrial weighing systems. Manual weighing approaches are labor-intensive and time-consuming, and they induce stress in both animals and handlers, which substantially limits their suitability for frequent and repeated measurements [8], [9]. Although industrial weighing systems provide accurate and reliable measurements, their high acquisition, installation, and maintenance costs restrict their adoption to specific production settings and hinder their widespread use for routine monitoring [10]. As a consequence, body weight records collected under real farming conditions often remain incomplete, posing significant challenges for the reliable modelling and prediction of individual growth trajectories.

To alleviate the limitations associated with direct live-weight measurements, a variety of indirect estimation methods based on morphometric characteristics have been investigated. Approaches relying on body dimensions obtained through manual measurements or image-based extraction have demonstrated their potential for estimating cattle live weight with reasonable accuracy [11]. More recently, machine learning techniques have been increasingly employed to exploit such indirect measurements for weight prediction, providing enhanced modelling flexibility compared to traditional regression-based approaches [12]. However, these methods often require additional data acquisition devices, controlled measurement conditions, or specialized expertise, which may restrict their practical applicability across heterogeneous farming environments.

Beyond morphometric-based approaches, data-driven methods relying directly on historical body weight records have gained increasing attention for livestock weight prediction. By learning from longitudinal observations, machine learning models are able to capture nonlinear relationships between age, growth dynamics, and body weight that are difficult to represent using classical parametric regression techniques. Comparative analysis have shown that such models can achieve improved predictive performance for cattle live-weight estimation when sufficient explanatory information is available, as illustrated in studies conducted on Hereford cows [13].

In parallel, matrix factorization approaches have been proposed to explicitly address the sparsity and irregularity of longitudinal weight datasets. By representing animal weight measurements as incomplete matrices indexed by individuals and time points, these techniques aim to reconstruct missing observations through latent representations learned from the available data. This formulation has been successfully applied to cattle growth modelling, enabling the estimation of individualized weight trajectories under limited measurement conditions [14]. The methodological foundations of this approach originate from recommender systems, where matrix factorization has proven effective for learning from sparse and incomplete data structures [15].

Despite the progress achieved in livestock weight prediction, several limitations remain in existing approaches. Many studies focus on a single modelling strategy or rely on additional measurement systems, which makes it difficult to assess the relative effectiveness of different predictive models when only a limited number of weight measurements per animal are available. Consequently, a systematic comparative evaluation of predictive models capable of handling sparse longitudinal weight data remains insufficiently explored.

Building on these developments, the present study conducts a comparative evaluation of predictive models for livestock body weight trajectory estimation under low-measurement conditions. Matrix factorization approaches and state-of-the-art ensemble learning methods—including CatBoost, ExtraTrees, LightGBM, and XGBoost—are systematically assessed using a dataset of Holstein cows. Model performance is evaluated using standard regression metrics, namely the coefficient of determination, root mean squared error, mean absolute error, and mean absolute percentage error. The objective is to identify predictive models that achieve high accuracy while remaining robust to data sparsity, thereby enabling cost-efficient and practically deployable solutions for precision livestock farming.

The main contributions of this study can be summarized as follows:

- A systematic comparison between a matrix factorization model and several state-of-the-art ensemble learning algorithms (XGBoost, LightGBM, CatBoost, and ExtraTrees) for livestock body weight trajectory prediction.
- An evaluation framework designed specifically for low-measurement conditions, where only a limited number of historical weight observations per animal are available.
- A robust experimental protocol based on five-fold cross-validation performed at the animal level to prevent data leakage and ensure reliable performance assessment.
- An interpretability analysis using SHAP to identify the most influential predictors of cattle body weight evolution and to better understand the contribution of growth-related features.
- Practical insights for precision livestock farming by demonstrating that accurate weight prediction can be achieved using only a small number of routine weight measurements collected under real farm conditions.

II. Related Works

Early studies on livestock body weight estimation mainly relied on regression-based approaches and parametric growth models to describe the relationship between age and body weight. Although these methods are generally simple and interpretable, they often fail to capture nonlinear growth patterns and inter-animal variability, particularly when measurements are sparse or irregular [5], [7].

To improve prediction accuracy, indirect approaches based on morphometric measurements have been widely investigated. Traditional body size measurements have long been used for livestock weight estimation, and more recent studies have extended this direction through digital image analysis and computer vision techniques [8], [11]. In particular, machine learning models combined with image-derived morphological traits have shown promising performance for cattle weight prediction [13], [19]. However, these approaches usually require additional sensing or image acquisition systems, as well as controlled acquisition conditions, which may limit their applicability in routine farming environments [20].

More recently, data-driven machine learning methods operating on livestock growth and body weight data have received increasing attention. Comparative studies have shown that modern machine learning algorithms can outperform conventional approaches in cattle weight prediction tasks. For example, a comparative analysis of machine learning algorithms for Hereford cows reported strong predictive performance for ensemble-based methods [12]. In parallel, recent research has highlighted the growing role of machine learning and deep learning for modelling cattle weight gain and identifying the factors influencing livestock growth dynamics [17]. A further study proposed a machine learning framework for mob-based cattle weight gain forecasting using Random Forest, Support Vector Regression, and Long Short-Term Memory models, and showed that Random Forest achieved the best performance when historical weight records were combined with environmental variables such as rainfall and temperature [18]. These studies confirm the increasing relevance of machine learning techniques for modelling cattle growth and improving predictive accuracy in livestock production systems.

To explicitly address the problem of sparse and incomplete longitudinal records, matrix factorization techniques have also been introduced for livestock growth modelling. By learning latent representations from incomplete weight matrices, these methods enable the reconstruction of missing observations and provide an effective strategy for modelling body weight trajectories under limited measurement conditions [14],[15].

Despite these advances, few studies have provided a comprehensive comparison between sparsity-oriented matrix factorization methods and state-of-the-art ensemble-based machine learning models under strictly low-measurement settings using only routine body weight records. Moreover, many recent approaches rely on additional morphometric, image-derived, or environmental variables [18], [20] which may improve prediction performance but reduce practical deployability in routine farm conditions. This limitation motivates the present comparative study, which evaluates matrix factorization and ensemble-based machine learning methods under explicitly controlled low-measurement scenarios using only historical body weight data.

III. Methodology

A. Data Collection

The dataset used in this study was collected from a private cattle farm located in the Taroudant region, Morocco. It consists of live body weight records from 2,686 Holstein cattle. All animals are male and were born and raised within the same farm.

Data collection was conducted under routine farm management conditions between 2016 and 2020. Live body weight was measured manually using a conventional weighing scale and recorded in kilograms. For each animal, five body weight measurements are available, recorded at different ages ranging from birth to 512 days. Because weight measurements were not acquired at fixed or regular time intervals, the dataset contains sparse and partially observed records. This structure reflects real farming conditions and is suitable for evaluating predictive models under low-measurement settings, without relying on dense temporal sampling or controlled experimental protocols.

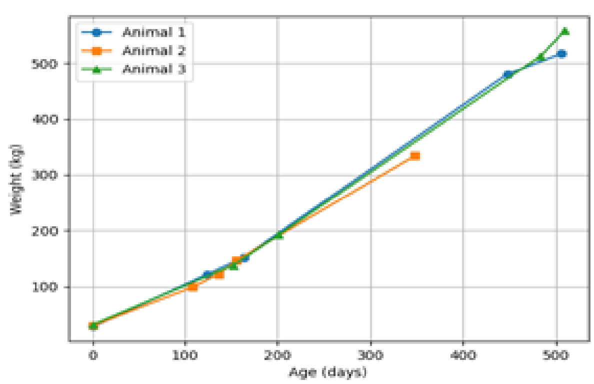

Fig. 1. Example growth curves of individual animals illustrating variations in trajectory shape, growth rate (slope), and initial age (at birth).

- Considerable variability in growth rates is observed among animals of identical sex and breed.

This variability substantially increases the complexity of the learning task. Fig. 1 illustrates representative weight curves highlighting the heterogeneity in growth patterns across animals. In addition to inter-individual differences, the sparsity of weight records further complicates modelling, as measurements are limited and irregular. Moreover, acquiring frequent weight measurements requires time and financial resources, which may not be feasible for small-scale farms.

B. Preprocessing

Prior to model development, a preprocessing step was carried out to ensure the quality and consistency of the dataset. The initial dataset comprised 10,672 animals collected under routine farm management conditions, with a variable number of body weight measurements ranging from 2 to 6 observations per animal and including different cattle breeds. To obtain a more homogeneous population for modelling, only animals belonging to the Holstein breed were retained for the present study. In addition, to ensure consistency in the experimental setup, only animals with five body weight measurements recorded at different ages during the growth period were considered in the final dataset used for model development.

To detect abnormal observations, an outlier detection procedure based on the z-score normalization method was applied using the Scikit-learn library [16]. The z-score measures how far an observation deviates from the mean of the dataset in terms of standard deviations and is defined as Eq. (1):

$$z = \frac{x-u}{s} \tag{1}$$

where, x represents the observed value, u denotes the mean of the dataset, and s corresponds to the standard deviation. Observations with a z-score greater than 3 or lower than −3 were considered outliers and removed from the dataset. This threshold makes it possible to eliminate extreme values while preserving the natural biological variability of cattle growth.

Descriptive statistics of the main variables in the final filtered dataset are summarized in Table I. The mean body weight was 255.48 kg, with a standard deviation of 196.94 kg, indicating substantial variability across different stages of animal development. Similarly, the average age was 183.16 days, with a standard deviation of 166.90 days, reflecting a wide distribution of ages in the dataset. The relatively high coefficients of variation (77.09% for body weight and 91.12% for age) indicate a heterogeneous dataset covering multiple growth phases of the animals.

TABLE I. Descriptive Statistics of the Dataset Variables

Parameters	Mean	SD	CV (%)
Body Weight (kg)	255.48	196.94	77.09
Age (days)	183.16	166.90	91.12

The dataset initially comprised 10,672 body weight measurements collected under routine farm management conditions. These measurements corresponded to 2,686 animals, with approximately four weight observations recorded for each animal at different ages. The mean body weight was 255.48 kg, with a standard deviation of 196.94 kg, indicating substantial variability in body weight. Similarly, the average age was 183.16 days, with a standard deviation of 166.90 days, reflecting a wide age distribution in the dataset. The relatively high coefficients of variation (77.09% for body weight and 91.12% for age) indicate that the dataset captures animals at different stages of growth, as summarized in Table I.

C. Feature Space

In this study, we adopt the feature space formulation proposed [14]. The model relies on vector representations for both animals and ages, where each animal a and age d is associated with a latent feature vector. These latent vectors define a shared embedding space that captures intrinsic growth-related patterns across individuals.

The representation enables the model to differentiate animals based on their implicit growth behavior while simultaneously encoding age-dependent variations in body weight. By projecting both entities into a common latent space, the approach models the interaction between animal-specific characteristics and temporal development. This formulation is particularly suitable for sparse datasets, as it allows the estimation of unobserved weight values through learned latent factors rather than relying solely on explicit measurements.

1) Animal feature vectors: Animals are described using features derived from their first m weight measurements, where m represents the number of historical measurements available for prediction. For an animal a with weight measurements taken at ascending ages, we have Eq. (2):

$$a = \{(a_{g_i}, lw_i)\}_{i=1}^{n} \tag{2}$$

where, a_{g_i} is the i-th age in days and lw_i is the i-th live weight in kilograms.

The feature representation varies depending on the value of m:

- For $m = 1$ (single measurement available): The reduced feature vector contains:
  - ag_1: age at the first weight measurement
  - Iw_1: first recorded live weight
  - $rwa_1 = Iw_1 / ag_1$: ratio of weight to age at first measurement
- For $m \geq 2$ (multiple measurements available): The extended feature vector includes:
  - ag_1 : age at the first weight measurement
  - ag_m: age at m-th measurement
  - $\overline{ag}m$: average age across the first m measurements
  - Iw_1: first recorded live weight
  - lw_m: m-th recorded live weight
  - $\overline{lw_m}$: average live weight across the first m measurements
  - $rwa_1 = Iw_1 / ag_1$: ratio of weight to age at first measurement
  - $a\,dg_m = (lw_m - Iw_1)/(ag_m - ag_1)$: average daily gain computed from the first m measurements

These features capture both the initial state of the animal when measurements begin and the growth dynamics observed in the available historical data. The average daily gain ($a\,dg_m$) is particularly important as it provides information about the animal's growth rate, which helps predict future weights. We are focused solely on males, and holstein cows. Therefore, we ignore sex and race in this study.

2) Age feature vector: Two different feature representations were considered in this study. we adopted the same polynomial age encoding used in the original formulation. Specifically, age was represented through a cubic polynomial expansion (age, age^2, age^3), where age denotes the number of days at which body weight is predicted. This representation enables the model to capture nonlinear growth trends within the latent factor framework.

In contrast, for the gradient boosting models, age was included as a single numerical variable $d =$ age. Polynomial expansion was not required, as tree-based ensemble methods inherently model nonlinear relationships and interaction effects through recursive partitioning. This distinction ensures that each modelling approach is evaluated under a feature representation consistent with its structural properties.

D. Machine Learning Algorithms

Five predictive models were evaluated for cattle weight prediction, including a matrix factorization approach based on the formulation proposed in [14] and four ensemble-based machine learning models: XGBoost, LightGBM, CatBoost, and ExtraTrees. The objective of this study is to compare the predictive performance of matrix factorization and ensemble-based learning methods under low-measurement conditions. The matrix factorization model was included to address data sparsity through latent representation learning, whereas the ensemble-based models were selected for their ability to capture nonlinear relationships in heterogeneous and limited tabular data. The ensemble-based models were implemented in Python using standard machine learning libraries, including Scikit-learn [16], while the matrix factorization model was implemented in Python following the formulation described in [14].

1) Matrix factorization: Matrix factorization was used in this study as a sparsity-aware approach for predicting cattle body weight trajectories under low-measurement conditions [14]. The problem was formulated as a matrix completion task in which animals correspond to rows, ages correspond to columns, and observed body weight measurements correspond to known matrix entries. The model learns two low-dimensional embedding matrices, W and V , which map animals and ages into a common latent space of dimension k. The predicted body weight for animal $a$at age $d$is computed as given below, where W_a and V_d represent the latent representations of the animal and age, respectively, as defined in Eq. (3):

$$\hat{g}(a,d) = \langle W_a, V_d \rangle \quad (3)$$

The matrix factorization model was trained using Stochastic Gradient Descent (SGD) to minimize the mean squared error with L_2 regularization. The hyperparameter configuration adopted in this study is summarized in Table II. The latent space dimension was set to $k = 10$, the learning rate decay parameter was fixed at $\gamma_s = 1.0$, and the regularization coefficient was defined as $\nu = 0.001$. The model was trained for 100 iterations, which provided stable convergence during training. Preliminary sensitivity analyses indicated that further adjustments of these hyperparameters did not lead to significant improvements in predictive performance.

TABLE II. HYPERPARAMETER CONFIGURATION OF THE MATRIX FACTORIZATION MODEL.

Parameter	Value	Description
k	10	Latent space dimension
γ_s	1.0	Learning rate decay parameter
ν	0.001	L2 regularization coefficient
Iterations	**100**	**Number of SGD training iterations**

2) Ensemble-based tree models: Ensemble-based tree learning methods construct predictive models by combining multiple decision trees in order to improve prediction accuracy and generalization. In this study, four tree-based ensemble algorithms were evaluated for cattle weight prediction under low-measurement conditions: XGBoost, LightGBM, CatBoost, and ExtraTrees. These models were selected because of their strong performance in regression tasks involving tabular data and their ability to capture complex nonlinear relationships.

TABLE III. HYPERPARAMETER CONFIGURATION OF THE GRADIENT BOOSTING MODELS.

Parameter	XGBoost	LightGBM	CatBoost
Boosting rounds	100	100	100
Maximum depth	6	-	6
Learning rate	0.05	0.05	0.05
Row sampling	0.8	0.8	0.8
Column sampling	0.9	0.9	-
num_leaves	-	31	-
bagging_freq	-	5	-

The detailed hyperparameter configuration of these models is reported in Table III.

XGBoost (Extreme Gradient Boosting) implements a regularized boosting framework that incorporates both L_1 and L_2 penalties to control model complexity. In this study, the model was configured with 100 estimators, a maximum tree depth of 6, a learning rate of 0.05, a row subsampling ratio of 0.8, and a column subsampling ratio of 0.9. These settings provide a balance between model flexibility and generalization by controlling tree depth and introducing stochastic sampling to mitigate overfitting.

LightGBM (Light Gradient Boosting Machine) adopts a leaf-wise tree growth strategy combined with histogram-based optimization, which improves computational efficiency while maintaining high predictive performance. The model was trained using 100 boosting rounds, a maximum of 31 leaves per tree, a learning rate of 0.05, a feature fraction of 0.9, a bagging fraction of 0.8, and a bagging frequency of 5 iterations.

CatBoost (Categorical Boosting) relies on ordered boosting and symmetric tree structures, which help reduce prediction bias and improve training stability. In this study, CatBoost was configured with 100 iterations, a maximum tree depth of 6, a learning rate of 0.05, and a subsampling ratio of 0.8.

TABLE IV. HYPERPARAMETER CONFIGURATION OF THE EXTRATREES MODEL.

Parameter	Value	Description
n_estimators	100	Number of trees
max_depth	None	Unlimited tree depth
min_samples_split	2	Minimum samples to split a node
min_samples_leaf	1	Minimum samples per leaf node

ExtraTrees (Extremely Randomized Trees) was included in this study as one of the evaluated tree-based ensemble models for cattle weight prediction under low-measurement conditions. The hyperparameter configuration of the model is summarized in Table IV. The model was implemented with 100 estimators and no explicit maximum depth constraint, allowing the trees to grow until purity. The minimum number of samples required to split an internal node was set to 2, and the minimum number of samples per leaf node was fixed at 1. This configuration preserves model flexibility while maintaining the model's ability to capture nonlinear relationships in heterogeneous tabular data.

E. Model Evaluation

To ensure methodological consistency with [14] and maintain a comparable evaluation framework, we adopted the same experimental protocol based on two key parameters: m_{train}, representing the number of measurements used for model training, and m_{test}, representing the number of measurements available at test time. In our experiments, both parameters were set equal ($m_{train} = m_{test} = m$) to guarantee that models were evaluated under the same information constraints used during training.

The value of m varies from 1 to 3, given that each animal is associated with a maximum of five recorded weight measurements. For a given configuration m, the first m measurements are used as input features, and the model is trained to predict the remaining future measurements. For example, when $m = 1$, the model uses only the first weight observation to predict subsequent measurements corresponding to positions 2 through 5. Similarly, when $m = 2$, the first two measurements are used as input, and the model is trained to predict the weights at positions 3, 4, and 5.

This evaluation strategy allows systematic analysis of model performance under progressively increasing information availability, while preserving a consistent and controlled experimental setup across all predictive approaches.

1) Data splitting and cross-validation: A five-fold cross-validation strategy was adopted to evaluate model performance. Data partitioning was carried out at the animal level rather than at the individual measurement level in order to prevent data leakage. This ensures that all records associated with a given animal are assigned entirely to either the training set or the testing set within each fold, thereby providing a realistic evaluation of the models' ability to generalize to unseen animals.

For each fold, the dataset was divided into 80% training data and 20% testing data. Stratified sampling based on animal identifiers was applied to preserve the distribution of animals across the subsets. To guarantee reproducibility and enable consistent comparisons between models, predefined random seeds (30, 50, 70, 85, and 90) were used during the data splitting process.

2) Evaluation metrics: Model performance was evaluated using three complementary regression metrics, each capturing a distinct aspect of predictive accuracy.

Mean Squared Error (MSE) was computed in the standardized feature space (after z-score normalization) in order to prevent scale effects from influencing error magnitude. It is defined as Eq. (4):

$$\mathrm{M\Sigma E} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{w}_i - w_i)^2 \tag{4}$$

where, $\widehat{w}_i$ and w_i denote the predicted and true standardized weights, respectively, and n represents the number of observations.

Mean Absolute Error (MAE) was calculated in kilograms to provide an interpretable measure of the average absolute deviation between predicted and actual weights in the original measurement scale [see Eq. (5)]:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} | \widehat{w}_i - w_i | \tag{5}$$

Mean Absolute Percentage Error (MAPE) expresses prediction error as a relative percentage of the true weight, allowing scale-independent comparison across animals with different body weight ranges [see Eq. (6)]:

$$\mathrm{MA\Pi E} = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{\widehat{w}_i - w_i}{w_i} \right| \times 100 \tag{6}$$

Together, these metrics provide a comprehensive evaluation framework by quantifying squared error magnitude (MSE), absolute deviation (MAE), and relative prediction accuracy (MAPE).

IV. Results and Discussion

A. Predictive Performance Under Low-Measurement Conditions

In this study, five predictive models were evaluated for cattle weight prediction: the Matrix Factorization (MF) approach proposed in [14] and four tree-based ensemble algorithms, namely XGBoost, LightGBM, CatBoost, and ExtraTrees [16]. Particular emphasis was placed on low-measurement conditions, defined as scenarios in which only a limited number of historical weight records per animal are available. This setting reflects practical livestock farming constraints, where frequent weight measurements are costly and difficult to obtain.

To simulate sparse-data conditions, the number of available measurements per animal (m) was restricted to 1, 2, and 3. These configurations make it possible to assess model robustness as the amount of historical information increases. For each value of m, a five-fold cross-validation procedure was performed at the animal level to prevent data leakage. Identical predefined random seeds were used across all models to ensure consistent data splits and a fair comparison between methods. Model performance was evaluated using the mean and standard deviation of MSE, MAE (kg), and MAPE (%).

The results are presented in Table V to Table VII for $m = 1$, $m = 2$, and $m = 3$, respectively. Under the most restrictive condition ($m = 1$), where only a single historical weight measurement is available per animal, the Matrix Factorization (MF) model exhibits substantially higher error values. This behaviour indicates a strong sensitivity to data sparsity under low-measurement conditions. Since the MF model relies on latent factor representations learned from the interaction between animals and ages, the limited amount of information available in this setting restricts its ability to estimate reliable latent factors.

In contrast, the tree-based ensemble models maintain stable and significantly lower error values, demonstrating greater robustness under low-measurement conditions. Models such as XGBoost, LightGBM, CatBoost, and ExtraTrees are able to capture nonlinear relationships between age and weight even when the available historical information is limited. Their recursive tree construction enables them to model complex variable interactions without requiring a large number of previous measurements.

TABLE V. Results of 5-Fold Cross-Validation for M = 1 Using the Most Efficient Algorithms [Mean and SD for MSE, MAE (kg), and MAPE (%)].

Model	MSE		MAE (kg)		MAPE (%)	
	Mean	SD	Mean	SD	Mean	SD
XGBoost	0.04794	0.00285	26.544	0.726	9.192	0.323
LightGBM	0.04815	0.00369	26.739	0.722	9.258	0.325
CatBoost	0.04908	0.00283	27.466	0.701	9.707	0.280
ExtraTrees	0.05940	0.00290	29.060	0.597	9.791	0.271
Matrix Factorization	0.09965	0.00241	44.016	1.042	19.436	0.688

Among the evaluated methods, XGBoost achieves the best overall predictive performance for $m = 1$ in Table V. This superior performance can be attributed to its regularized boosting framework, which combines multiple decision trees while controlling model complexity. These regularization mechanisms help prevent overfitting and improve generalization, making XGBoost particularly effective in sparse-data scenarios. The other ensemble-based models show comparable accuracy levels, confirming the overall robustness of tree-based ensemble approaches under extreme measurement constraints.

TABLE VI. RESULTS OF 5-FOLD CROSS-VALIDATION FOR M = 2 USING THE MOST EFFICIENT ALGORITHMS [MEAN AND SD FOR MSE, MAE (KG), AND MAPE (%)].

Model	*MSE*		*MAE (kg)*		*MAPE (%)*	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
ExtraTrees	0.06425	0.00121	29.734	0.395	8.307	0.163
XGBoost	0.06056	0.00399	29.831	0.776	8.470	0.290
LightGBM	0.06095	0.00408	30.100	0.711	8.691	0.258
CatBoost	0.06448	0.00427	31.616	1.014	9.259	0.379
Matrix Factorization	0.09651	0.01061	42.769	2.747	14.566	1.314

For $m = 2$ in Table VI, the performance gap between the Matrix Factorization model and the tree-based ensemble models remains evident. Although the availability of two measurements improves the predictive capability of all models, the ensemble-based approaches continue to achieve substantially lower error values across MSE, MAE, and MAPE. This improvement can be explained by the additional historical information provided by the second measurement, which allows the models to better capture the early growth dynamics of each animal.

In this configuration, the ExtraTrees model slightly outperforms XGBoost, achieving the lowest mean MAPE and MAE values. This result suggests that the additional randomness introduced by ExtraTrees during the split selection process helps improve model generalization when a moderate amount of historical information is available. By constructing multiple highly diversified trees, ExtraTrees can better adapt to heterogeneous growth patterns across animals.

Overall, these results confirm that tree-based ensemble methods maintain strong predictive accuracy and generalization capability under limited measurement settings. Even when only two historical measurements are available, these models are able to capture meaningful relationships between age and body weight, leading to reliable predictions.

For $m = 3$ in Table VII, the increase in the number of available historical measurements leads to a noticeable improvement in the performance of the Matrix Factorization (MF) model, highlighting its dependence on a larger amount of historical information. With three observed measurements per animal, the MF model is able to estimate more reliable latent representations of the interaction between animals and ages.

TABLE VII. RESULTS OF 5-FOLD CROSS-VALIDATION FOR M = 3 USING THE MOST EFFICIENT ALGORITHMS [MEAN AND SD FOR MSE, MAE (KG), AND MAPE (%)].

Model	*MSE*		*MAE (kg)*		*MAPE (%)*	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
ExtraTrees	0.08963	0.00499	32.591	0.383	7.734	0.118
XGBoost	0.09037	0.00540	33.285	0.183	7.955	0.078
LightGBM	0.08907	0.00526	33.279	0.264	8.004	0.076
CatBoost	0.09942	0.00643	35.383	0.894	8.687	0.305
Matrix Factorization	0.10986	0.01263	39.616	2.327	10.384	0.798

However, despite this improvement, the MF model does not surpass the tree-based ensemble models in any of the evaluated metrics. The ensemble-based methods continue to provide lower prediction errors and more stable results across folds. Under this configuration, the ExtraTrees model again achieves the best overall predictive accuracy, followed closely by XGBoost and LightGBM.

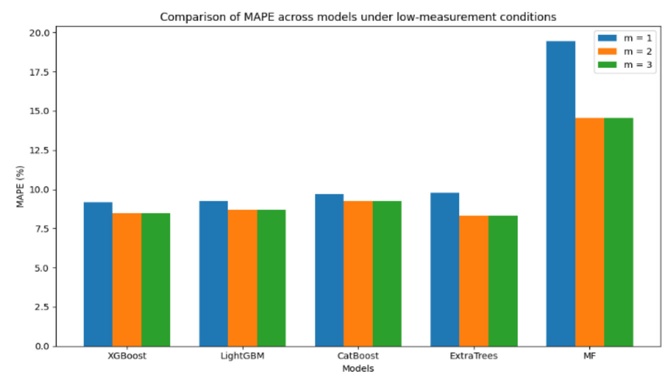

Fig. 2. Comparison of MAPE values across the evaluated models under low-measurement conditions ($m = 1,2,3$). Ensemble-based models consistently achieve lower prediction errors than Matrix Factorization.

Fig. 2 shows the MAPE comparison of the evaluated models for $m = 1, 2,$ and 3. The tree-based ensemble models consistently produce lower percentage errors than Matrix Factorization across all scenarios. XGBoost achieves the best performance for $m = 1$, whereas ExtraTrees performs best for $m = 2$ and $m = 3$. Although Matrix Factorization improves when more measurements are available, it remains less accurate than the ensemble-based models.

The reduced variability observed for the ensemble-based models, reflected by lower standard deviation (SD) values, further indicates greater stability and robustness across cross-validation folds. These results suggest that tree-based ensemble models are better suited to modelling heterogeneous growth patterns, as they can capture nonlinear relationships between age and body weight while remaining robust under limited-measurement conditions.

Overall, the results demonstrate that although Matrix Factorization benefits from additional measurements, tree-based ensemble models consistently achieve higher predictive accuracy and greater robustness across all low-measurement scenarios. These findings confirm their suitability for practical livestock applications where frequent body weight measurements are difficult to obtain.

B. Model Interpretation Using SHAP Analysis

To complement the quantitative evaluation reported in Table V to Table VII (for m = 1, 2, and 3, respectively), a SHAP (SHapley Additive exPlanations) analysis was conducted to investigate the contribution of input features to the predictions of the best-performing ensemble models under different measurement configurations.

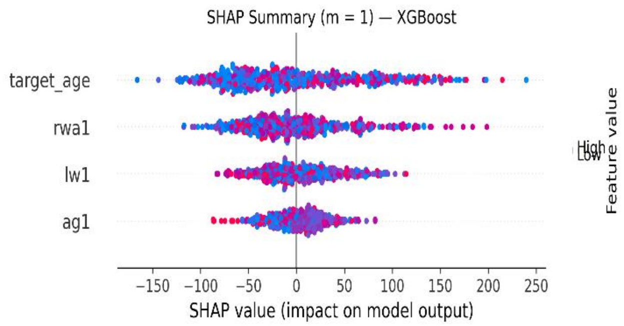

Fig. 3. Feature contribution analysis using SHAP for m = 1 (XGBoost model).

Fig. 3 presents the SHAP summary plot for the XGBoost model when only one historical measurement is available (m = 1). Under this extreme low-measurement condition, the model relies primarily on the initial weight measurement and the age at the first observation. Since no trajectory information is available, the prediction mainly depends on baseline physiological indicators describing the animal's initial condition. The SHAP values show that the first recorded weight has the largest influence on the predicted future weight, while the age variable provides complementary contextual information about the stage of growth.

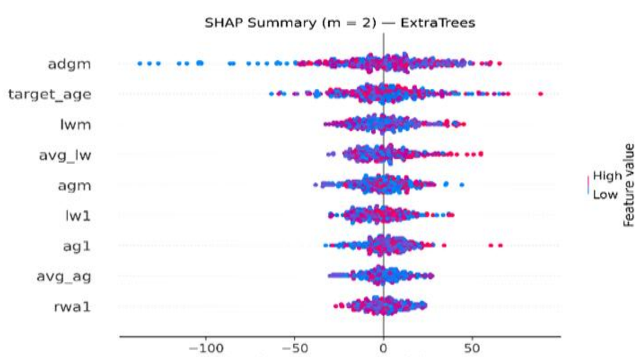

Fig. 4. Feature contribution analysis using SHAP for m = 2 (ExtraTrees model).

Fig. 4 illustrates the SHAP summary plot obtained for the ExtraTrees model when two measurements are available (m = 2). With the addition of a second observation, the model can start capturing the early growth dynamics of each animal. In this configuration, trajectory-related features such as the average weight and the average daily gain become more influential. The SHAP analysis indicates that growth rate indicators contribute significantly to the prediction process, allowing the model to better estimate the future evolution of body weight.

Fig. 5 presents the SHAP summary plot for the ExtraTrees model when three measurements are available (m = 3). In this configuration, the model benefits from richer historical information, enabling a more accurate representation of the animal's growth trajectory. The SHAP results highlight the dominant contribution of growth-related variables, particularly average daily gain and aggregated weight indicators. These variables provide direct information about the animal’s growth velocity, which explains the improved predictive performance observed in Table VII.

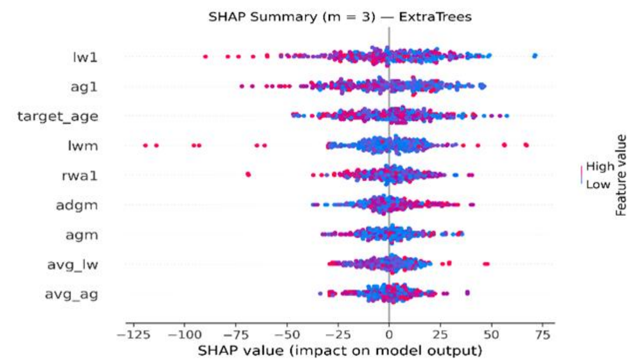

Fig. 5. Feature contribution analysis using SHAP for m = 3 (ExtraTrees model).

To complement the quantitative evaluation reported in Table V to Table VII, a SHAP (SHapley Additive exPlanations) analysis was conducted to interpret the contribution of input features to the predictions of the best-performing ensemble models. Fig. 3 shows the SHAP summary plot for m = 1 (XGBoost), where the prediction mainly depends on the initial weight measurement due to the absence of trajectory information. Fig. 4 presents the SHAP analysis for m = 2 (ExtraTrees), where growth-related features such as average weight and early growth indicators become more influential. Finally, Fig. 5 illustrates the SHAP results for m = 3 (ExtraTrees), highlighting the dominant contribution of trajectory-based variables, particularly those describing growth dynamics. These results confirm that as the number of available measurements increases, the models rely more on growth-related features to improve prediction accuracy.

1) Global feature importance across measurement configurations: Across all configurations, the SHAP analysis reveals that dynamic growth indicators play a dominant role in prediction accuracy. For m = 2 and m = 3, in Fig. 4 and Fig. 5, Average Daily Gain (ADG) consistently emerges as the most influential predictor, followed by the first recorded weight (lw1), the average weight (avg_lw), and the target age. This ranking is consistent with the superior predictive performance reported in Tables V and VI, where the ensemble models achieve lower MSE, MAE (kg), and MAPE (%) values.

The dominance of ADG indicates that growth velocity provides more predictive information than isolated static measurements. From a biological perspective, the early growth rate reflects metabolic efficiency and genetic growth potential, which explains its strong contribution within tree-based ensemble models.

2) Interpretation under extreme sparsity (m = 1): Under extreme low-measurement conditions (m = 1; see Table V and Fig. 3), the model lacks sufficient trajectory information and must rely primarily on initial physiological indicators, including the first recorded weight (lw1), the age at the first measurement (ag1), and the weight-to-age ratio (rwa1). In this configuration, the SHAP values indicate a stronger dependence on variables describing the animal's initial body condition.

Since only one historical measurement is available, the model cannot infer the growth dynamics of each animal and instead extrapolates future weight based on baseline physiological characteristics. This limitation explains the relatively higher prediction errors observed in Table V, particularly for the Matrix Factorization model, which struggles under extreme sparsity compared with gradient boosting approaches.

3) Evolution of feature contributions with increasing m: As the number of available measurements increases (m = 2 and m = 3), the SHAP results Fig. 4 and Fig. 5 show a clear shift toward trajectory-based variables, particularly Average Daily Gain (ADG), the average weight (avg_lw), and aggregated growth indicators. This transition corresponds to the performance improvements observed in Table VI to Table VII, where both the error magnitude (MAE and MAPE) and the variability across folds (SD) decrease.

The ensemble models are able to exploit nonlinear interactions between growth rate and age more effectively when additional historical information becomes available. As a result, the models achieve improved predictive stability, reduced fold-to-fold variance, and greater robustness to sparsity in the weight measurements.

4) Biological and methodological implications: The SHAP interpretation confirms that the superiority of gradient boosting models, as demonstrated in Table V to Table VII, is not merely statistical but also biologically coherent. Under low-measurement conditions, the models prioritize dynamic growth descriptors rather than isolated static measurements. This behavior validates both the relevance of the feature formulation inspired by [14] and the effectiveness of ensemble learning methods for handling sparse longitudinal livestock data.

Overall, the interpretability analysis strengthens the scientific credibility of the proposed framework by demonstrating that the most influential variables correspond to biologically meaningful growth indicators. In addition, the ensemble models progressively adapt as the number of available measurements increases, allowing them to capture growth dynamics more effectively. These findings indicate that accurate livestock weight prediction can be achieved even under low-measurement conditions, which is particularly important for small-scale or resource-constrained farming systems where frequent data collection is difficult to implement.

V. Conclusion

This study demonstrates that gradient boosting models consistently outperform matrix factorization techniques for livestock body weight prediction under low-measurement conditions. While the performance of matrix factorization improves when additional historical measurements become available, ensemble tree-based methods—particularly XGBoost and ExtraTrees—maintain superior predictive accuracy even when only a small number of weight observations per animal are available. These results indicate that reliable livestock weight prediction can be achieved without relying on dense longitudinal records, image-based sensing systems, or extensive morphometric measurements.

The main contribution of this work lies in the systematic comparison of matrix factorization and state-of-the-art gradient boosting models under explicitly controlled low-measurement scenarios (m = 1, 2, and 3). Unlike approaches that require complex sensing infrastructures, the proposed framework relies exclusively on sparse body weight records collected under routine farm management conditions. This design demonstrates that ensemble learning models remain robust when operating with incomplete and irregular livestock data.

Beyond the methodological contribution, the findings have direct practical implications for precision livestock farming. Accurate prediction of cattle body weight using only a limited number of measurements can assist farmers and farm managers in monitoring growth trajectories, optimizing feeding strategies, and determining appropriate market or slaughter timing. By reducing the need for frequent manual weighing or costly measurement technologies, the proposed approach offers a practical and cost-efficient solution for supporting daily farm management decisions.

Such predictive tools are particularly beneficial for small- and medium-scale farms or production systems with limited technological resources. Integrating machine learning–based prediction models into farm management platforms could enable more efficient monitoring of herd performance, facilitate data-driven decision-making, and improve overall productivity within precision livestock farming systems.

Future research will focus on extending the proposed framework by incorporating additional environmental and management-related variables, such as feeding regime, temperature, and humidity, to further improve predictive performance and provide a deeper understanding of the factors influencing cattle growth dynamics in real-world farming environments.

References

[1] G. B. West, J. H. Brown23, and B. J. Enquist, "A general model for ontogenetic growth," 2001. [Online]. Available: www.nature.com

[2] H. C. Freetly, L. A. Kuehn, R. M. Thallman, and W. M. Snelling, "Heritability and genetic correlations of feed intake, body weight gain, residual gain, and residual feed intake of beef cattle as heifers and cows," J. Anim. Sci., vol. 98, no. 1, Jan. 2020, doi: 10.1093/jas/skz394.

[3] E. Frigo, C. D. Dechow, O. Pedron, and B. G. Cassell, "The genetic relationship of body weight and early-lactation health disorders in two

experimental herds," J. Dairy Sci., vol. 93, no. 3, pp. 1184–1192, Mar. 2010, doi: 10.3168/jds.2009-2313.

[4] J. Dí Ez et al., "Artificial intelligence techniques point out differences in classification performance between light and standard bovine carcasses," 2003. [Online]. Available: www.elsevier.com/locate/meatsci

[5] J. Alonso, A. Bahamonde, A. Villa, and Á. R. Castañón, "Morphological assessment of beef cattle according to carcass value," Livest. Sci., vol. 107, no. 2–3, pp. 265–273, Apr. 2007, doi: 10.1016/j.livsci.2006.09.027.

[6] J. Alonso, Á. R. Castañón, and A. Bahamonde, "Support Vector Regression to predict carcass weight in beef cattle in advance of the slaughter," Comput. Electron. Agric., vol. 91, pp. 116–120, 2013, doi: 10.1016/j.compag.2012.08.009.

[7] J. Alonso, A. Villa, and A. Bahamonde, "Improved estimation of bovine weight trajectories using Support Vector Machine Classification," Comput. Electron. Agric., vol. 110, pp. 36–41, 2015, doi: 10.1016/j.compag.2014.10.001.

[8] C. Enevoldsen and T. Kristensen, "Estimation of Body Weight from Body Size Measurements and Body Condition Scores in Dairy Cows," J. Dairy Sci., vol. 80, no. 9, pp. 1988–1995, 1997, doi: 10.3168/jds.S0022-0302(97)76142-3.

[9] F. Coopman, S. De Smet, H. Laevens, A. Van Zeveren, and L. Duchateau, "Live weight assessment based on easily accessible morphometric characteristics in the double-muscled Belgian Blue beef breed," Livest. Sci., vol. 125, no. 2–3, pp. 318–322, Nov. 2009, doi: 10.1016/j.livsci.2009.05.005.

[10] Z. Wang, S. Shadpour, E. Chan, V. Rotondo, K. M. Wood, and D. Tulpan, "ASAS-NANP SYMPOSIUM: Applications of machine learning for livestock body weight prediction from digital images," 2021, Oxford University Press. doi: 10.1093/jas/skab022.

[11] S. Tasdemir, A. Urkmez, and S. Inal, "Determination of body measurements on the Holstein cows using digital image analysis and estimation of live weight with regression analysis," Comput. Electron. Agric., vol. 76, no. 2, pp. 189–197, May 2011, doi: 10.1016/j.compag.2011.02.001.

[12] A. Ruchay et al., "Comparative analysis of machine learning algorithms for predicting live weight of Hereford cows," Comput. Electron. Agric., vol. 195, no. February, p. 106837, 2022, doi: 10.1016/j.compag.2022.106837.

[13] V. A. de Moraes Weber et al., "Prediction of Girolando cattle weight by means of body measurements extracted from images," Revista Brasileira de Zootecnia, vol. 49, 2020, doi: 10.37496/RBZ4920190110.

[14] J. Alonso, J. Díez, O. Luaces, and A. Bahamonde, "Original papers A new method to learn growth curves of beef cattle using a factorization approach," Comput. Electron. Agric., vol. 151, no. October 2017, pp. 77–83, 2018, doi: 10.1016/j.compag.2018.05.033.

[15] Yehuda Koren, "COVER FE ATURE," 2009.

[16] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A.,Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machinelearning in Python. Journal of Machine Learning Research 12, 2825–2830.

[17] M. R. H. Hossain, R. Islam, S. R. McGrath, M. Z. Islam, and D. Lamb, "Learning-based estimation of cattle weight gain and its influencing factors," Comput. Electron. Agric., vol. 231, no. February, p. 110033, 2025, doi: 10.1016/j.compag.2025.110033.

[18] M. R. H. Hossain, M. R. Islam, S. R. McGrath, M. Z. Islam, and D. Lamb, "Mob-based cattle weight gain forecasting using ML models," Dec. 01, 2025, Elsevier B.V. doi: 10.1016/j.atech.2025.101428.

[19] De Oliveira, F. M., Ferraz, P. F. P., Ferraz, G. A. e. S., Pereira, M. N., Barbari, M., & Rossi, G. (2025). Prediction of Body Mass of Dairy Cattle Using Machine Learning Algorithms Applied to Morphological Characteristics. Animals, 15(7). https://doi.org/10.3390/ani15071054

[20] Ma, W., Qi, X., Sun, Y., Gao, R., Ding, L., Wang, R., Peng, C., Zhang, J., Wu, J., Xu, Z., Li, M., Zhao, H., Huang, S., & Li, Q. (2024). Computer Vision-Based Measurement Techniques for Livestock Body Dimension and Weight: A Review. In Agriculture (Switzerland) (Vol. 14, Number 2). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/agriculture14020306